2008

Degenerate Codon Mutagenesis and Cloning of G-Protein Coupled Receptor Gene Fusion

Michael J. Sofolarides
Seton Hall University

Degenerate Codon Mutagenesis and Cloning of G-Protein Coupled Receptor Gene Fusions

by

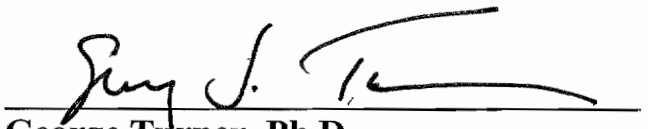
Michael J. Sofolarides

M. S. Thesis

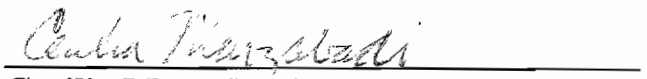
Submitted to the Department of Chemistry and Biochemistry of Seton Hall University in partial fulfillment of the requirements for the degree of Master of Science.

We certify that we have read this thesis and that in our opinion it is sufficient in scientific scope and quality as a thesis for the degree of Master of Science in Chemistry.

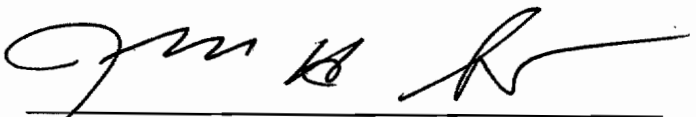
Approved by:



George Turner, Ph.D.
Mentor



Cecilia Marzabadi, Ph.D.
Committee Chair



Nicholas Snow, Ph.D.
Committee Member
and
Chair of the Department of Chemistry and Biochemistry

TABLE OF CONTENTS

DEDICATION

For my family: past, present and future

Acknowledgement

I would like to express my thanks to the following people who have given a great deal of their time, consideration and support during the preparation of this degree:

Dr. George Turner - a mentor and a friend
Dr. Nicholas Snow - department chair and committee member
Dr. Cecelia Marzabadi - committee member
Dr. Steve Kelty - graduate chair and advisor
Helen Kubowicz - department administrator
Jason Guzman - researcher in the Turner group
Maged Darwish - researcher in the Turner group
Dr. Dong Xiao - supervisor, Schering-Plough Research Inst.
Dr. Anandan Palani - supervisor, Schering-Plough Research Inst.
Dr. Robert Aslanian - supervisor, Schering-Plough Research Inst.
Stephanie Ciesla - colleague and friend

LIST OF FIGURES

LIST OF TABLES

Introduction

How do we as humans respond to our environment? We act on questions like: "Is it cold? Is something burning? Is that a baby crying?" How do the cells in our fingers, nose, ears, eyes, and brain interpret environmental information such as heat, pain, odorants, light, and hormones? These extracellular signals are recognized by specific "receptor" proteins (G-Protein Coupled Receptors, GPCR) embedded in a cell's outer-membrane. The receptor initiates a cellular response by physical interaction with intracellular signal transfer proteins (G-proteins; **Figure 1**). GPCR:G-protein interactions are essential for normal function of nearly all cell types. The work presented in this thesis represents a necessary step toward the physical analysis of GPCR:G-protein interactions; the generation of gene fusions that will allow production and purification of GPCR fusion receptor proteins.

G-Protein Coupled Receptors are extremely sensitive. GPCR are naturally produced at very low levels in individual cells, so as not to overwhelm all other cellular functions. Indeed, over-activity of receptors can lead to cell death or cancer. However, the low abundance of GPCR in cells has severely crippled scientific analyses of their mechanisms of action. Despite heroic effort from numerous laboratories, obtaining receptor proteins in the quantities required for their purification and physical analysis remains a forefront problem [**1,2**]. A logical approach to create a system for the over-production of receptor proteins is to exploit a cell type or organism that naturally produces large quantities of a membrane protein

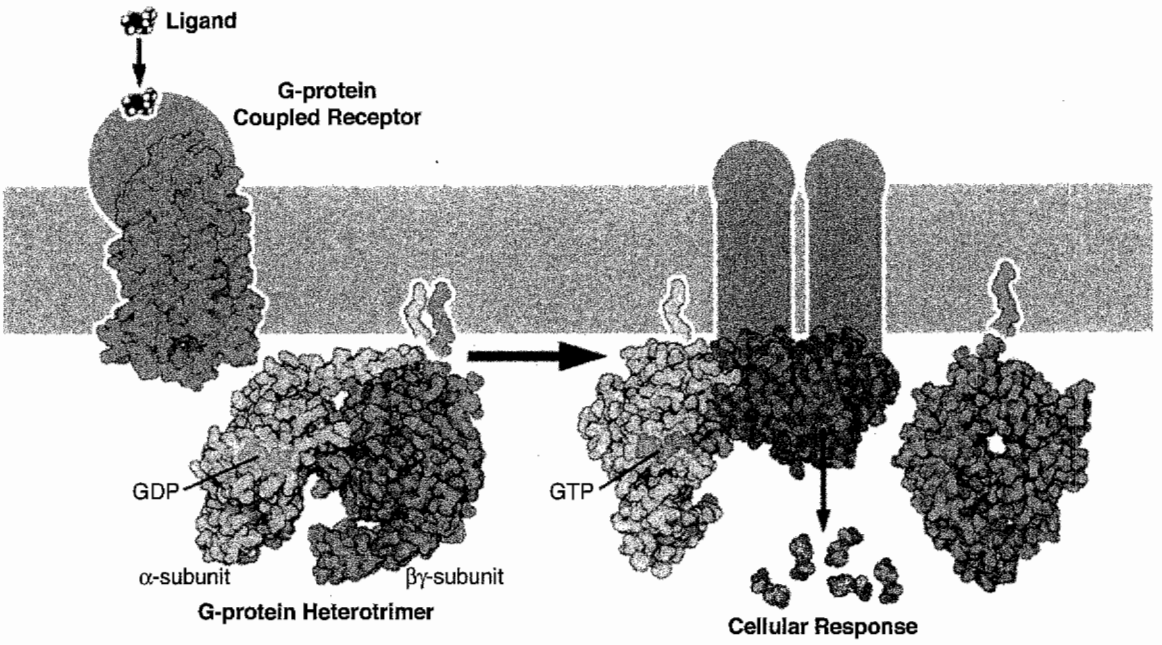


Figure 1: GPCR Signal Transduction

The extracellular ligand couples to the G-protein coupled receptor. This "activated" receptor causes an exchange of guanosine diphosphate (GDP) for guanosine triphosphate (GTP) activating the G_{α}-subunit shown in yellow. This "activated" G_{α}-subunit dissociates from the $G_{\beta\gamma}$-subunits, shown in blue and green, respectively, and couples to the secondary protein shown in purple. Coupling of the activated G_{α}-subunit to this secondary protein causes a cellular response, shown in red. In this case, the $G_{\beta\gamma}$ dimer is not involved in signaling. The result is the ligand's "signal" being transduced across the cell membrane, shown in gray, by the G-protein coupled receptor and its associated G-proteins.

Adapted from an illustration by David S. Goodsell of The Scripps Research Institute.

and genetically engineer it for general usage. Our laboratory has pursued such a strategy. The aim of our work is to use this new expression system to obtain receptor proteins in the quantities useful for chemical and physical analysis of GPCR:G-protein interactions. Our strategy exploits the molecular machinery responsible for the naturally abundant membrane protein bacteriorhodopsin (bR) in the archaeal organism *Halobacterium salinarum (H. salinarum)*. Herein, we demonstrate the use of this system to make bR:GPCR fusion proteins in the quantities required to map the structure of GPCR:G-protein interaction sites. This is done by replacement of the third intracellular loop of bR with the topological equivalent from the Mam2 receptor, a GPCR found in the fission yeast *Schizosaccharomyces pombe* (*S. pombe*). The loop region was replaced using degenerate codon mutagenesis, a mutagenic strategy that employs *in vitro* gene evolution to direct maximal expression of receptor fusion proteins.

G-Protein Coupled Receptors

G-Protein Coupled Receptors are a diverse family of membrane proteins comprising the largest single category of sensory receptors. It is well established that GPCR are essential for normal cellular function [**3**]. At least 1000 subtypes have been identified and at least 200 have known functions. Defective receptors or GPCR:G-protein interactions lead to diseases such as breast cancer, congestive heart failure, schizophrenia, high blood-pressure, allergic reactions, stroke, cognitive disorders, and blindness [**4-6**]. Approximately 60 percent of all prescribed drugs are now known to target GPCR and account

for 200 billion dollars in sales per year [7]. Studies focused on GPCR functional mechanisms are critical to advances in modern molecular science and human health.

Despite their obvious importance and the amount of work focused on GPCR function, our understanding is limited because the structure and structural changes coupled to receptor activation are unknown. The only two GPCR whose structures are known at the atomic level are the visual pigment bovine rhodopsin [**8-11**] and, recently the β2 adrenergic receptor [**12**]. Characterization of all other GPCR structures has therefore relied on sequence homology alignments [**13-14**], predictive secondary structural algorithms [**15**] and mutagenesis [**16-18**]. Genomic and proteomic analyses indicate that all GPCR possess an extracellular amino-terminus (N-terminus), seven alpha-helical transmembrane spanning domains (7TM) and an intracellular carboxyl-terminus (C-terminus) [**14**]. The transmembrane domains are bridged by "loop regions" located on the extracellular (e1-3) and intracellular (i1-3) surface of the receptor (**Figure 2**). The intracellular loop regions communicate the activation state of the receptor to associated G-proteins.

GPCR are stimulated by interaction with an extracellular signal (e.g. ligand: a drug, hormone, light, or a peptide/protein). The signaling event is "transduced" across the cellular membrane via ligand-dependent activation of the receptor; a process termed "signal transduction". For many pharmaceutical targets, much is known about the ligand-receptor affinities and the specific

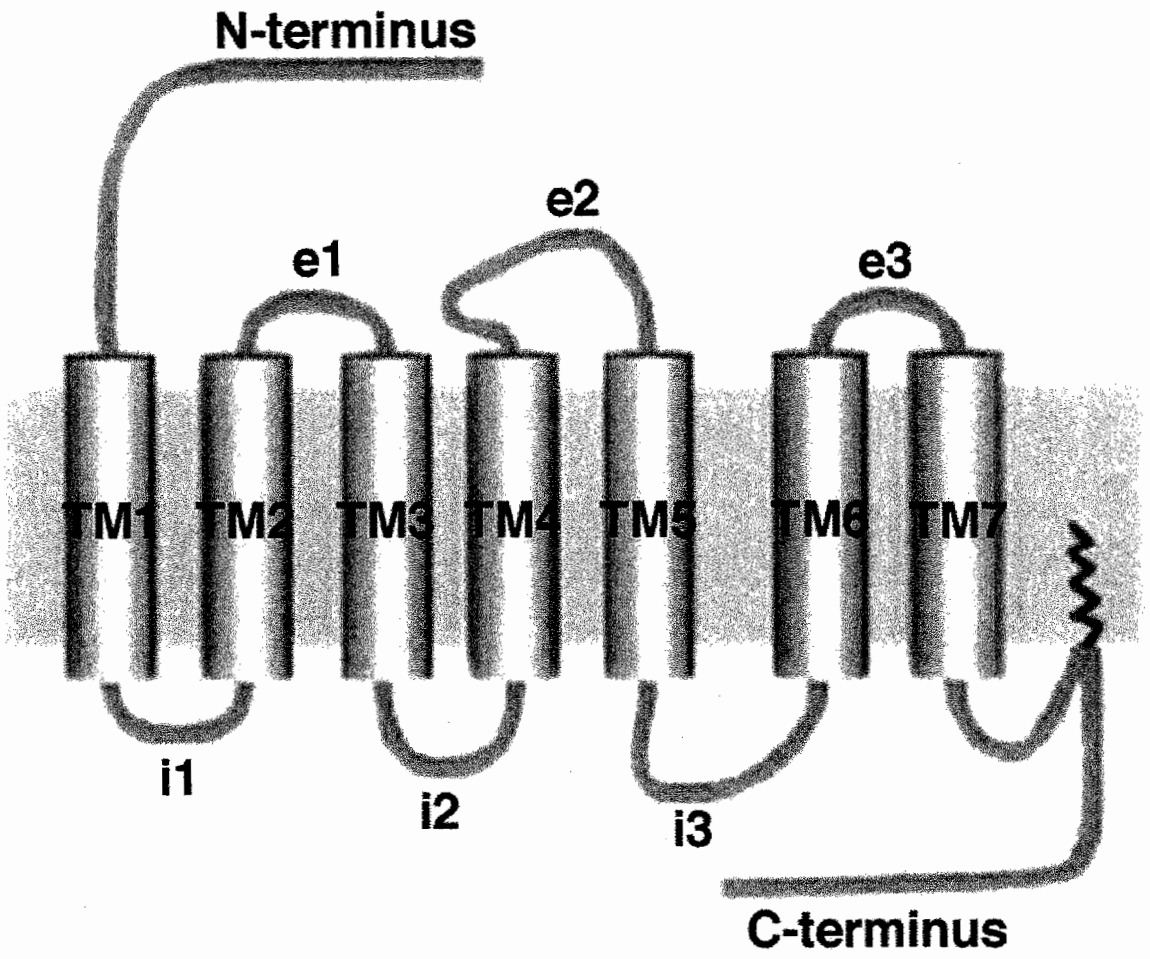


Figure 2: Two-dimensional Structure and Nomenclature of GPCR

The structure of a GPCR begins outside the cell with the amino-terminus (N-terminus), and ends inside the cell with the carboxyl-terminus (C-terminus). The silver cylinders, representing the 7-transmembrane (TM) spanning domains characteristic of GPCR, are numbered TM1 through TM7 and shown "spanning" the cell membrane, shown in blue. These domains are connected via extramembranous loop regions that are either extracellular (e1, e2, e3) or intracellular (i1, i2, i3).

events downstream of receptor activation [**19-21**]. Molecular engineering has been particularly useful in identifying amino acids that contribute to ligand-binding affinities [**22**]. However, we are ignorant of how ligand-binding generates an "active" receptor. Likely, the receptor undergoes a change in shape, or structural transition, in response to ligand-binding. Ligand-induced receptor structural changes are communicated to specific G-proteins in contact with the receptor *inside* the cell membrane [**23,24**]. Mutagenesis has been useful for identifying GPCR amino acids contributing to G-protein interaction and selectivity [**3,25,26**]. The GPCR:G-protein interaction is likely mediated through the intracellular loop regions (i1-3) and C-terminus. G-proteins are typically heterotrimeric proteins consisting of α, β, and γ subunits. GPCR interaction is mediated through the C-terminal end of the G-protein α subunit [**3,21,27**]. GPCR activation causes exchange of guanosine diphosphate (GDP) for guanosine triphosphate (GTP) on the G-protein α subunit resulting in dissociation from the receptor and the $G_{\beta\gamma}$ dimer. Both G_{α} and $G_{\beta\gamma}$ moieties are considered active and either or both can elicit unique cellular responses, such as enzyme and ion channel activity, or gene expression [**27**]. A critical step in G-protein activation must involve discrimination of liganded and unliganded receptor structures, at least at the level of the intracellular loop regions.

The three-dimensional crystallographic structures of bacteriorhodopsin (bR) and bovine rhodopsin have served as productive structural and functional models for GPCR. Detailed spectroscopic analysis has revealed that these

light receptors transiently populate multiple structural states during their photosynthetic and visual mechanisms, respectively [13,28-31]. By analogy, GPCR activation may proceed through several ligand-dependent structural changes [19,32,33]. Distinct receptor conformations can differentially regulate G-protein activities, which may account for the variety of biological responses observed for activation of single receptor types [34,35]. The existence of multiple receptor conformations would be directly testable through biochemical and biophysical methods (e.g. crystallography and spectroscopy). However, these studies have been severely hampered by naturally low GPCR expression levels and the inability of available protein expression systems to generate large quantities of GPCR [23,36,37]. It is a goal of our work to generate bR:GPCR fusion proteins in quantities that will allow structural and spectroscopic investigation of G-protein activation.

GPCR/G-Protein Coupling and Selectivity

The functional roles that the GPCR loop regions play in receptor conformation and G-protein selectivity and activation are understood at a primitive level. Few studies have supported a direct role for the i1 loop in G-protein recognition [3,38]. For a wide range of GPCR subclasses, the i1 loop is short and a nearly constant size of five to seven amino acid residues [39]. This conserved feature has been taken in support of a structural, rather than signaling role for loop i1. The i2 loop has been implicated in contributing to GPCR:G-protein coupling and selectivity. For example, the i2 loop of the thrombin

receptor is specific for coupling to the $G_q/_{11}$ family of G-proteins. When engineered into other GPCR, the i2 loop allowed coupling of the chimeric receptor to the $G_q/_{11}$ G-proteins [**40**]. Genetic engineering of the vasopressin (V) receptor demonstrated that replacing the i2 loop of the G_s-coupled V_2 receptor with that of the $G_{q/11}$-coupled V_{1a} receptor yielded a receptor that coupled to $G_{q/11}$ proteins [**41**]. The i3 loop has been widely reported as essential for GPCR/G-protein selectivity through use of mutagenesis and fusion proteins. Mutagenesis of the serotonin, or 5-hydroxytryptamine receptor, 5-HT_6, indicates that the i3 loop is responsible for coupling to the G_s family of G-proteins [**42**]. Also, deletion mutagenesis studies on the N-terminal (Ni3) and C-terminal (Ci3) segments of the i3 loop have been shown to play a key role in G-protein coupling in the α2 adrenoreceptor [**43**]. In addition, chimeric receptors containing i3 loop swaps have resulted in G-protein selectivity in non-GPCR scaffolds [**23,44**]. The C-terminus, or i4 region, can also play a role in G-protein coupling selectivity. In many receptors, including the neurokinin A and avian β-adrenergic, deletion mutagenesis was used to demonstrate that the membrane proximal amino acids (8-16 amino acids) interact with G-proteins, while the distal amino acids simply contain phosphorylation sites that regulate receptor desensitization [**45,46**]. The i4 of the Ste2p receptor in the yeast *Saccharomyces cerevisiae (S. cerevisiae)* has been shown to be involved in receptor desensitization, however not critical in coupling with endogenous G-proteins [**47,48**].

Mam2 Receptor

Yeast are among the most useful systems for the analysis of GPCR due to their robust genetics, short generation time and relative simplicity [**49-52**]. Reporter strains of *S. cerevisiae* and *S. pombe* have been constructed and shown to be useful in the investigation of functional mechanisms of human GPCR [**50,53**].

The haploid fission yeast *S. pombe* contains cells of two mating types, h+ (P) and h- (M), each producing its own pheromone, P-factor and M-factor, respectively. The pheromone binding receptor on the h- cells is Mam2, a type 4 or "D" GPCR family member [**53**]. Mam2:P-factor interaction initiates the mating response by activation and dissociation of the G_α subunit, Gpa1 [**53**]. Gpa1 functions as a monomeric G-protein; no $G_{\beta\gamma}$ subunits have been shown to be associated with Mam2-dependent signaling [**55**]. *S. pombe* is known to express a single G_β subunit and a single G_γ subunit, however this dimer interacts exclusively with Gpa2, in a separate glucose-sensing pathway [**56**]. Cloning and sequencing of the *mam2* gene locus identified a 348 amino acid open reading frame coding for a 39 kilodalton receptor protein [**57**]. There have been no published investigations of the Mam2 receptor structure.

Mating Response

S. pombe h+ cells secret P-factor which binds to Mam2 receptors in h- cell types. Stimulation of Mam2 initiates signal transduction through activation of Gpa1, leading to the mating response [**56**]. Gpa1 acts through the switch-

like protein, Ras1, initiating transcriptional and morphological responses. Transcriptional activation results in the programmed breakdown of P-factor (negative regulation) through a MAP kinase cascade [**55**]. Morphologically, the cells elongate to form a "shmoo" in preparation for fission [**56**]. The full details of events downstream of Gpa1 activation have yet to be determined as are the critical receptor interactions leading to Mam2:Gpa1 coupling. However, it is known that removal of the C-terminus from a homologous pheromone receptor in the yeast *S. cerevisiae* (Ste2p) does not inhibit signal transduction [**47,48**]. Therefore, the C-terminus is not involved in Gpa1 activation. In addition, conservation of loop length and limited mutagenesis of amino acid residues in i3 loops of pheromone receptor homologues implicate a key role for i3 loops in Gpa1 activation [**58**]. In summary, we hypothesize that the i3 loop is critical for the pheromone response pathway in *S. pombe*, and have chosen the Mam2 i3 loop as a test of the bR:GPCR fusion receptor strategy.

Halobacterium salinarum

Our preferred host organism for over-expression of GPCR and bR:Mam2 fusion receptors is the archaeon *Halobacterium salinarum (H. salinarum)*. *H. salinarum* is a rod shaped prokaryote found in evaporating salt ponds across the globe. Under intense light and/or anaerobic conditions, *H. salinarum* expresses the photosynthetic light receptor, bacteriorhodopsin (bR) at extremely high densities. bR can occupy up to 80 percent of the cell membrane in *H. salinarum* and can be purified in yields up to 15-30 milligrams per liter of cell culture [**59,60**]. Under

aerobic conditions *H. salinarum's* viability does not depend on bR accumulation. The cell membrane may therefore be available for expression of non-native receptor proteins. As a 7TM protein, bR is a structural homologue of the GPCR. Therefore it is logical to exploit *H. salinarum* for the attempted over expression of GPCR.

This goal is tractable owing to many system advantages. As a member of the archaea, *H. salinarum* shares many traits with both prokaryotes and eukaryotes, which may facilitate the expression of proteins from either domain [**61,62**]. *H. salinarum* does not possess G-protein signal transduction components; therefore, GPCR over-expression is not anticipated to affect cell physiology and viability [**63**]. Vectors for expression of genetically engineered bR and GPCR, and methods for gene transfer have been developed [**63**]. Techniques for *H. salinarum* culturing and manipulation are easy and similar to those used for *Escherichia coli (E. coli)* [**59,60**]. Quantitative cell fractionation is accomplished by hypotonic lysis and membrane fractions are purified by centrifugation. Finally, the sucrose density ultracentrifugation procedure used to purify bR has been used to isolate GPCR and bR:GPCR fusion receptors expressed in *H. salinarum* cell membranes [**64-66**].

The *bacterio-opsin* gene and Bacteriorhodopsin

Since its discovery in 1967, bacteriorhodopsin (bR), the simplest known light-driven proton pump, has served as the prototypical membrane protein due to its simplicity, stability and high levels of expression in *H. salinarum*

[67]. bR is a 1:1 assembly of the chromophore retinal and the 248 amino acid bacterio-ospin protein. Retinal is covalently bound by a Schiff-base linkage to lysine 216 in the seventh transmembrane helix of bacterio-opsin [68]. Although bR is a membrane protein, it is not a GPCR. The extremely high-levels of bR expression is quite unlike that of a GPCR. This high level of expression, along with the chromophoric properties of retinal, produce *H. salinarum* cultures with an intense purple color leading to an efficient visual assay of expression [60] (**Figure 3**). Although bR is structurally homologous to a GPCR, the extramembranous regions are typically shorter in length than GPCR, likely due to the lack of extracellular ligand binding and coupling to intracellular signaling components (**Figure 4**).

The remarkable levels of bR that accumulate are attributable to the efficient expression of the *bacterio-opsin* (*bop*) gene in *H. salinarum*. The *bop* gene has been cloned, sequenced and the transcriptional promoter, putative regulatory factor binding sites, and transcriptional terminator have been identified [69,70,71]. In addition, there is molecular machinery dedicated to the insertion of bR into the cell membrane. The first 11 amino acids function as a signal sequence, coupling *bop* gene translation to membrane insertion [72]. Some of the most dramatic membrane protein successes result from homologous expression of bR in *H. salinarum*. Under the control of the *bop* gene promoter, numerous bR mutants and low abundance homologues, halorhodopsin (HR) and sensory-rhodopsin (SR), have been expressed in yields of 1-15 milligrams per liter of *H. salinarum* cell culture [63,65,66,73-77]. Exploiting

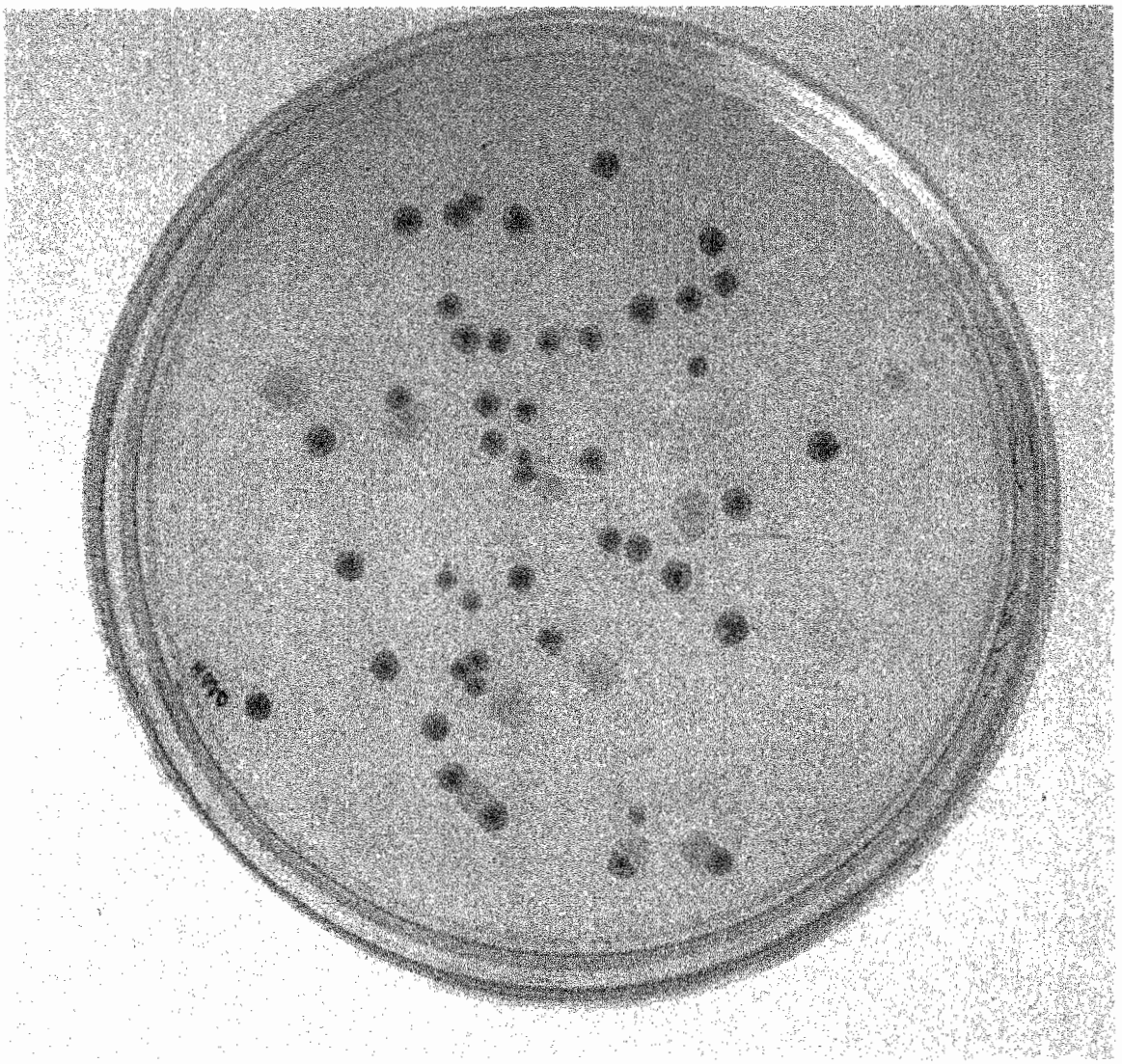


Figure 3: *H. salinarum Cell Colonies on Solid Media*

Colonies of *H. salinarum* strains L33, white colonies, and ET1001, purple colonies, grown on non-selective media. The purple membranes result from high levels of expression of the membrane protein bacteriorhodopsin.

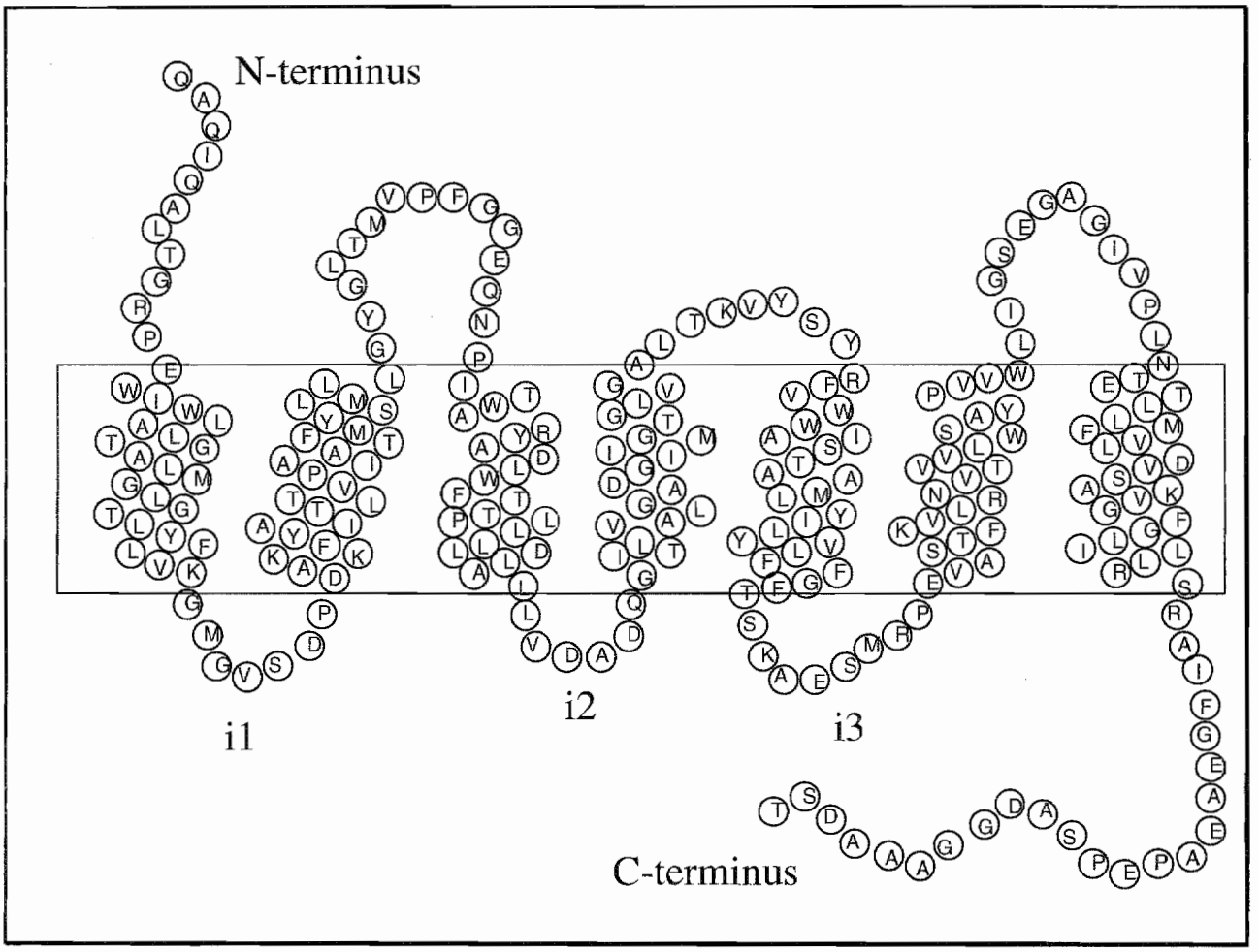


Figure 4: Structure of Bacteriorhodopsin

The structure of bacteriorhodopsin (bR) is shown with amino acids represented by their one-letter abbreviations. The seven transmembrane spanning helical domains (7TM), homologous to those of a GPCR, are indicated inside the cell membrane, represented by the black box. Intracellular loop regions i1, i2, and i3 are colored blue. The orientation of the N-terminus and C-terminus are arbitrary.

the potential of the *bop* gene to create a prokaryotic GPCR expression system is a logical extension of these efforts. It is noted that the levels of expression of these bR homologues, and even some bR variants containing point mutations, do not match those routinely achieved for the wild-type *bop* gene; an indication that perturbation of the *bop* gene sequence can influence gene expression.

A receptor-specific expression system based on the *bop* gene is under development in the Turner laboratory. Prototype cloning vectors (pENDS, for "ends" of the *bop* gene) linking the coding region of GPCR genes (*gpcr*) to *bop* gene transcription and translation control elements have been constructed and tested [**63**]. The amino acids comprising the *bop* amino-terminal signal sequence, which couple translation to membrane insertion, are included in all gene fusions. The pENDS vectors were engineered to contain carboxyl-terminal coding sequences that "tag" the heterologous gene products. The tags are short peptide sequences: one is an epitope from the Influenza Hemagglutinin protein (HA), the other is six consecutive histidines (6His). These tags facilitate immunogenic detection and purification of expressed receptor proteins [**78**]. The pENDS vectors also generate DNA fragments with cloning sites allowing transfer of the fused *bop:gpcr* genes into *H. salinarum* expression vectors (designated pHex, for *Halobacterium* expression vectors) [**63,65**].

GPCR Expression Tests

This expression system has been tested for its ability to generate eukaryotic receptor proteins. The Ste2 pheromone

receptor was expressed at low levels in *H. salinarum* [**63,65**]. Ste2 formed a high-density membrane fraction that was isolated by the same sucrose density centrifugation protocol used to purify bR. Bovine rhodopsin (Rho) was also expressed at low levels, as detected by Western analysis [**63**]. The human Angiotensin receptor (AT_{1A}) was expressed as four different bR:AT_{1A} gene fusions. While protein accumulation was not indicated, full length mRNA was detected by RNA gel blot analysis [**63**]. Functional $\alpha 2_B$-adrenergic receptors were expressed at levels as high as 40 picomoles per milligram of total cell protein [**66**]. In addition, another group has expressed low levels of functional β2-adrenergic receptors in *H. salinarum* by a similar strategy [**79**]. While functional full-length GPCR can be expressed in *H. salinarum*, the yields obtained, to date, are orders of magnitudes below that for the wild-type gene. The lower levels of expression that are achieved when *bop* gene coding regions are replaced with those of non-native receptors further suggests that there is something unanticipated and special about the coding region of the *bop* gene.

GPCR Fusions

While expressing full-length GPCR at levels comparable to bR remains a longer term goal [**63**], we consider that the *bop* gene expression system may be productive for expression of bR:GPCR fusion receptor proteins in the shorter term. As discussed previously, specific GPCR domains are essential (and transferable) for G-protein coupling and selectivity. Therefore, while pursuing methods to maximize receptor expression we have pursued a more modest and

experimentally tractable goal of expressing GPCR functional domains in the bR structural scaffold. Since the structural transitions of bR have been extensively characterized [**67,68**], bR:GPCR fusion proteins that couple with G-proteins, will provide insight regarding receptor conformation and G-protein coupling and selectivity.

As a first test of this gene fusion strategy a 26 kilodalton *E. coli* soluble protein (the catalytic subunit of Aspartate Transcarbamylase, not a GPCR) was expressed by fusing it, in frame, to the bR carboxyl-terminus. The fusion protein contained the bR 7TM helical-bundle. The construction was successfully expressed (3 milligrams per liter of cell culture), was purple, and formed a membrane fraction that was isolated by the identical sucrose density centrifugation protocol used to purify bR crystalline lattices. Electron diffraction verified native quaternary structure for both halves of the fusion, demonstrating that highly expressed and properly folded non-archaeal proteins can be produced [**50**].

In another study, 89 amino acids of the third intracellular loop region (i3) of the rat $\alpha 2_B$-adrenergic receptor were inserted into the bR i3 loop domain. The fusion protein was successfully expressed in *H. salinarum* at 0.5 milligrams per liter of cell culture and purified by the same sucrose density protocol used to isolate wild-type bR. The fusion protein bound retinal as evidenced by light- and dark-adapted ground-state visible absorbance spectra. Therefore, large coding region insertions did not destroy the bR functional fold and sufficient protein was expressed and purified to spectrally demonstrate function as a light-

driven proton pump and for crystallographic analysis [80].

While sufficient to characterize the fusion protein structure and function, levels of expression of the *bop* gene fusions (as measured by either mRNA or protein accumulation) were low and observed to vary depending on the *bop* gene fusion used. Indeed, the amount of mRNA that accumulated did not correlate with the extent of the *bop* gene coding region used in the fusion. This indicated that nucleotide sequences within non-native gene coding regions can unpredictably affect expression levels. We hypothesize that the "global" structure of the wild-type (wt) *bop* mRNA contributes to high levels of mRNA accumulation and the efficiency of translation. Inspection of the *bop* mRNA sequence helps to rationalize that the coding region contributes to gene expression. The wt *bop* mRNA is 831 nucleotides, 786 (94%) of which are coding region [69]. The *bop* mRNA is essentially leaderless; it contains 2 untranslated bases upstream of the codon for the initiating methionine. There are only 40 noncoding bases between the translation stop codon and mapped 3' end of the major transcript [81]. If nucleotides other than these downstream noncoding sequences contribute to *bop* mRNA accumulation and translation, they exist in the coding region. Since our strategy replaces the *bop* gene coding region with coding regions of non-native genes, sequence determinants that normally contribute to *bop* gene expression will be perturbed. We anticipate that identification of *bop* gene sequence-dependent determinants, and the development of methods to transfer them into *gpcr* gene coding regions, will be critical for maximizing the expression potential of this system.

Tags

Manipulation of carboxyl-terminal tag sequences revealed that important sequence-dependent determinants of over-expression exist within the *bop* gene coding region. The nine amino acid HA peptide and, separately, six sequential histidines were engineered to replace the terminal aspartic acid. The wt *bop* gene translational stop and transcriptional termination sequences were retained in all constructions. Expression of bR:HA resulted in a 40-fold reduction in mRNA and protein accumulation [**66**], while expression of bR:6His resulted in a 4-fold reduction in mRNA and protein accumulation. The two different sequences reduced levels of expression, but to different extents. This result demonstrates that specific nucleotide, and/or amino acid, sequences can critically impact expression.

In a separate experiment a DNA restriction site, ClaI, was introduced by mutating only two nucleotides immediately downstream of the *bop* gene translation stop sequence. Combining the ClaI site with the wild-type or bR:HA had a small positive effect on expression levels. The combination of the ClaI restriction site and bR:6His resulted in a near complete loss of mRNA and protein. Since the ClaI site was introduced downstream of the translation stop, the codons for the 6His tag were unchanged (recall that bR:6His without the ClaI site resulted in mRNA and bR accumulation at near wild-type levels). Taken in combination, these results demonstrate the extreme influence that nucleotide sequences (and not amino acid) in this region can have on *bop* gene expression.

Our results are similar to other observations that show carboxyl-terminal tag sequences can reduce expression; the effect was attributed to transcriptional perturbation [**82**]. Interestingly, a similar linkage between the carboxyl-terminal coding region and angiotensin type-1a receptor expression has been observed in mammalian cells [**83**].

To understand the basis of the sequence dependence observed, the carboxyl-terminal tagged *bop* genes were probed by semi-random mutagenesis (Turner and Guzman, unpublished). The degeneracy inherent in codon usage was exploited to construct large numbers of carboxyl-terminal coding sequences, where the amino acid sequence was not altered, and tested their effect on expression. The *bop* gene was chosen as a test case since the chromophoric properties of bR provide a simple cellular screen of expression. Colonies with high levels of bR expression are intensely purple, while colonies with low levels of bR expression are white. All possible 6His:ClaI and HA:ClaI codon combinations were cloned *en masse* into the *bop* gene and expressed (64 and roughly 10^6 sequences, respectively). 10% of the resulting mutants expressed high-levels of tagged bR. Recall the original bR 6His:ClaI construction did not accumulate any bR and that HA:ClaI expression was 10% of wt bR. These studies reveal that critical determinants of expression exist in the C-terminal *bop* gene coding region and these can be productively manipulated by a degenerate codon usage strategy.

Is the observed nucleotide sequence dependence on *bop* gene expression unique to the C-terminal coding region? To address this question we performed a similar degenerate

codon cassette analysis near the amino-terminal (N-terminal) coding region (Turner, unpublished). All possible synonymous codons were introduced into the coding region for amino acids 12 through 18. Amino acids 1-11 comprise the bR signal sequence. Since these amino acids are essential for co-translational membrane insertion they were not included in the degenerate codon mutagenesis analysis. The degenerate codon bank resulted in 3900 possible *bop* gene sequences, each coding for the wild-type N-terminal amino acids. The mutated *bop* gene sequences were transformed *en masse* into *H. salinarum*. *bop* and bR mRNA levels of accumulation were measured and the DNA sequences were determined for 84 unique transformants. The results indicate that *bop* gene expression is dependent on the coding sequence in this region also. A linear correspondence between *bop* mRNA levels and bR accumulation was observed and the *H. salinarum* G:C nucleotide bias was not obligatory. Structural analysis **[mFold, 82]** of the first 80 nucleotides in mutant *bop* sequences revealed no local structural determinants that correlated with expression phenotypes.

Taken in sum these studies reveal that:

1) bR:fusion proteins can be expressed at high-levels.
2) bR structure and function are retained in i3 and i4 fusion constructions.
3) Specific nucleotide sequences within the *bop* gene coding region are critical for high-level *bop* gene mRNA and bR accumulation.
4) Levels of *bop* gene mRNA and bR accumulation are tightly coupled.
5) Critical determinants of expression can be manipulated

by the degenerate codon mutagenesis strategy.

6) The sequence dependent determinants are not rare events. Approximately 10% of the degenerate codon mutagenesis constructions were highly expressed and over-producers were obtained.

The experimental work accomplished for this thesis makes use of the degenerate codon mutagenesis strategy to generate *bop:gpcr* gene fusions that will express at high-level in *H. salinarum*. The i3 loop of the *S. pombe* Mam2 receptor was chosen due to its small size and essential function as a critical component in G-protein coupling and selectivity. By selecting small functional domains as targets of the mutagenesis and expression we have simplified our experimental effort. We demonstrate that a pool of approximately 10^4 sequences coding for this bR:Mam2 i3 loop fusion protein have been engineered into bR cloning vectors (pENDs) and *H. salinarum* expression vectors (pHex). The work in this thesis describes seventy one individual clones from the pool that were isolated, purified, and sequenced. Thirty four clones and, separately, the entire pool of 10^4 sequences were subcloned into pHex vectors. The bR:Mam2 fusion constructions are currently being transformed in the *H. salinarum* strain L33 (bR^-) and their expression phenotypes will be characterized. The studies reported here are the necessary first steps to acquire the quantities of purified GPCR receptor mimics that will allow chemical and physical analysis of the initial stages of eukaryotic signal transduction.

MATERIALS AND METHODS

Bacterial Strains

The *Escherichia coli* K12 strain used was DH5α (*F*$^-$, *recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17* (r^-_k, m^+_k), *supE44*, l^-).

Media and Growth Conditions

All salts and chemicals were reagent grade from standard biochemical supply houses. Bacteriological peptone was from Oxoid (Unipath LTD., Hampshire, England); yeast extract tryptone, Difco peptone, and bacto-agar were from Difco Laboratories (Detroit, MI). Complex *E. coli* medium was yeast extract tryptone.

Reagents

Reagents were obtained from standard molecular biological supply houses. T4 DNA ligase and various restriction endonucleases were purchased from New England Biolabs (Beverly, MA) or from Boehringer Mannheim (Indianapolis, IN). Custom oligo-deoxynucleotide primers were purchased from Oligomer OY (Helsinki, FI). Oligo-directed mutagenesis was performed with the QuikChange II Site-Directed Mutagenesis Kit (Strategene, La Jolla, CA). Electrophoresis grade agarose was from FMC Corporation (Rockland, ME).

Protein and Gene Sequences of Mam2

The Mam2 receptor has been cloned and sequenced [**57**], however the structure of the Mam2 receptor is unknown. As part of this study we performed and compared several hydropathy predictions and homology sequence alignments to generate a putative Mam2 structural prediction (**Figure 5**) and membrane associated topology (**Figure 6**). This information can be used in conjunction with a putatively assigned structure to design construction of our bR:Mam2 fusion protein.

Mam2 Secondary Sequence Prediction

We compared several structural predictions to assign putative structure to the Mam2 receptor which can then be used to assign a putative Mam2 i3 loop amino acid sequence. This represents a more rational structural prediction through comparison of several structural predictions as opposed to designing our fusion protein experiment based on only one prediction.

Amino Acid Hydropathy Calculations

Engelman *et al* developed the Goldman, Engelman, Steitz (GES) hydrophobicity scale, based on experimental and theoretical considerations about how well each amino acid would enter the lipid bilayer from an aqueous environment [**84**]. A structural prediction of Mam2 was made by applying this hydrophobicity scale to the Mam2 amino acid sequence and the 7TM domains were putatively assigned (**Fig. 5, line 1**).

Kyte and Doolittle developed a hydropathy scale which is based on the hydrophobic and hydrophilic properties of the 20 amino

1) MRQPWWKDFTIPDASAIIHQNITIVSIVGEIEVPVSTIDAYERDRLLTGMTLSAQLALGV
2) MRQPWWKDFTIPDASAIIHQNITIVSIVGEIEVPVSTIDAYERDRLLTGMTLSAQLALGV
3) MRQPWWKDFTIPDASAIIHQNITIVSIVGEIEVPVSTIDAYERDRLLTGMTLSAQLALGV
4) MRQPWWKDFTIPDASAIIHQNITIVSIVGEIEVPVSTIDAYERDRLLTGMTLSAQLALGV
5) MRQPWWKDFTIPDASAIIHQNITIVSIVGEIEVPVSTIDAYERDRLLTGMTLSAQLALGV

1) LTILMVCLLSSSEKRKHPVFVFNSASIVAMCLRAILNIVTICSNSYSILVNYGFILNMVH
2) LTILMVCLLSSSEKRKHPVFVFNSASIVAMCLRAILNIVTICSNSYSILVNYGFILNMVH
3) LTILMVCLLSSSEKRKHPVFVFNSASIVAMCLRAILNIVTICSNSYSILVNYGFILNMVH
4) LTILMVCLLSSSEKRKHPVFVFNSASIVAMCLRAILNIVTICSNSYSILVNYGFILNMVH
5) LTILMVCLLSSSEKRKHPVFVFNSASIVAMCLRAILNIVTICSNSYSILVNYGFILNMVH

1) MYVHVFNILILLLAPVIIFTAEMSMMIQVRIICAHDRKTQRIMTVISACLTVLVLAFWIT
2) MYVHVFNILILLLAPVIIFTAEMSMMIQVRIICAHDRKTQRIMTVISACLTVLVLAFWIT
3) MYVHVFNILILLLAPVIIFTAEMSMMIQVRIICAHDRKTQRIMTVISACLTVLVLAFWIT
4) MYVHVFNILILLLAPVIIFTAEMSMMIQVRIICAHDRKTQRIMTVISACLTVLVLAFWIT
5) MYVHVFNILILLLAPVIIFTAEMSMMIQVRIICAHDRKTQRIMTVISACLTVLVLAFWIT

1) NMCQQIQYLLWLTPLSSKTIVGYSWPYFIAKILFAFSIIFHSGVFSY**KLFRAILIRKKIG**
2) NMCQQIQYLLWLTPLSSKTIVGYSWPYFIAKILFAFSIIFHSGVF**SYKLFRAILIRKKIG**
3) NMCQQIQYLLWLTPLSSKTIVGYSWPYFIAKILFAFSIIFHSGVFSYKL**FRAILIRKKIG**
4) NMCQQIQYLLWLTPLSSKTIVGYSWPYFIAKILFAFSIIFHSG**VFSYKLFRAILIRKKIG**
5) NMCQQIQYLLWLTPLSSKTIVGYSWPYFIAKILFAFSIIFHSGVFSY**KLFRAILIRKKIG**

1) **QFPFGP**MQCILVISCQCLIVPATFTIIDSFIHTYDGFSSMTQCLLIISLPLSSLWASSTA
2) **QFPFGP**MQCILVISCQCLIVPATFTIIDSFIHTYDGFSSMTQCLLIISLPLSSLWASSTA
3) **QFPFGPMQC**ILVISCQCLIVPATFTIIDSFIHTYDGFSSMTQCLLIISLPLSSLWASSTA
4) **QFP**FGPMQCILVISCQCLIVPATFTIIDSFIHTYDGFSSMTQCLLIISLPLSSLWASSTA
5) **QFPFGPMQCIL**VISCQCLIVPATFTIIDSFIHTYDGFSSMTQCLLIISLPLSSLWASSTA

1) LKLQSMKTSSAQGETTEVSIRVDRTFDIKHTPSDDYSISDESETKKWT
2) LKLQSMKTSSAQGETTEVSIRVDRTFDIKHTPSDDYSISDESETKKWT
3) LKLQSMKTSSAQGETTEVSIRVDRTFDIKHTPSDDYSISDESETKKWT
4) LKLQSMKTSSAQGETTEVSIRVDRTFDIKHTPSDDYSISDESETKKWT
5) LKLQSMKTSSAQGETTEVSIRVDRTFDIKHTPSDDYSISDESETKKWT

Figure 5: The Mam2 Secondary Structural Predictions

Five identical, overlaid amino acid sequences of the Mam2 receptor are shown. The amino acids are denoted using their one letter abbreviations. Transmembrane (TM) spanning domains are color coded or underlined and

extramembranous regions are in black. The first and second structural predictions were made by hydropathy calculations and the third and fourth made by sequence homology comparisons as follows:

Line 1: Hydropathy calculations as developed by Engelman *et al.* [**86**]

Line 2: Hydropathy calculations as developed by Kyte and Doolittle [**15**]

Line 3: Sequence homology with known primary structures in database [**87**]

Line 4: Sequence homology with known primary structures in database (www.genedb.org)

Line 5: Consensus of predictions as described in text

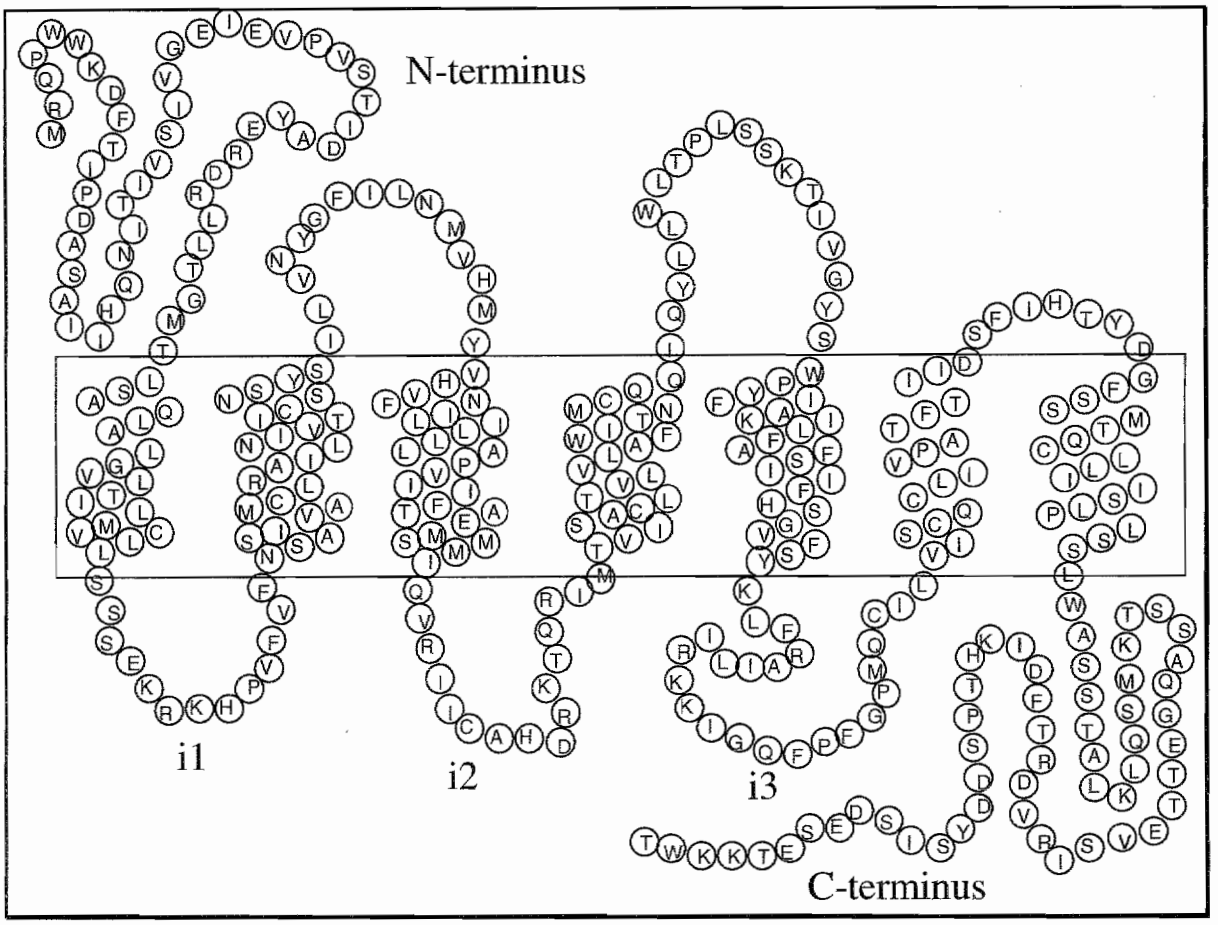


Figure 6: Assigned Structure of the Mam2 Receptor

Two-dimensional representation of Mam2 based on structural assignments made in Figure 5. Amino acids are labeled using their one-letter abbreviations. Putative transmembrane (TM) domains are present inside the cell membrane, represented by the black box. The intracellular domains (loops i1, i2, i3, and C-term) are shown in blue and are considered to be G-protein interacting domains. N-terminus and C-terminus folding pattern is arbitrary.

acids found in proteins [**15**]. This method was applied to the Mam2 amino acid sequence to putatively assign the 7TM domains of the Mam2 receptor (**Fig. 5, line 2**).

Amino Acid Homology Alignments

Rost *et al.* developed an algorithm that considers hydropathy calculations and then compares results against structures of known membrane proteins. For 89% of the test proteins used to form this algorithm, all transmembrane domains were predicted correctly [**85**]. The Mam2 sequence was evaluated with this approach and the 7TM domains were assigned (**Fig. 5, line 3**).

The GeneDB project is a core part of the Sanger Institute Pathogen Sequencing Unit's (PSU) activities and assigns 7TM topology to a queried amino acid sequence based on comparison with similar amino acid sequences of known structures within their database (**www.genedb.org**). A structural prediction of the Mam2 receptor was obtained using this sequence homology comparison and 7TM domains assigned (**Fig. 5, line 4**).

Each prediction method was compared side-by-side to identify Mam2 TM domains. Once the TM domains were assigned, the loop regions were identified and used to develop our fusion protein strategy and experimental design. The final structural assignment was done by empirical adjustment based on scientific intellect according to the following criteria: transmembrane domain amino acid agreement between all structural predictions, amino acid conservation and similarity as compared to the Ste2 receptor of *S. cerevisiae* [**58**], limited published mutagenesis of the Mam2 receptor, and future experimental design (**Figure 5, line 5**). Considering future experimental design, a lone cysteine residue (C249) was

observed in the area proximal to the intracellular interface of TM6 and the carboxyl end of the i3 region. We chose to include this residue in our structural assignment of the Mam2 i3 loop as it represents the only cysteine in the i3 loop and will be useful in future G-protein interaction studies by potentially facilitating sulhydryl labeling and protection studies. Based on our structural assignment of the Mam2, a two-dimensional depiction of membrane associated topology was constructed including assignment of the i3 loop amino acid sequence (**Fig. 6**).

Mam2 i3 Loop Design via Degenerate Codon Mutagenesis

The amino acid secondary structure prediction (**Figure 5, 6**) guided construction of the DCM Mam2 i3 loop (**Figure 7**). The oligonucleotides (oligos) for the coding and noncoding strands of the DCM Mam2 i3 loop region are shown (**Table I: 1, 2**). When annealed, the oligos generated upstream StyI and downstream SphI DNA restriction sites to facilitate cloning. Codons at the 5′ and 3′ ends of the oligos were chosen from abundant *H. salinarum* codons (**Table II**) [**86**] and were not made degenerate to facilitate annealing of the complementary strands. The *H. salinarum* codon bias is approximately 70 percent GC and the most abundant codons reflect that bias at the wobble base position (**Table II**). Since previous work indicated that about 10 percent of the DCM nucleotide sequences afforded high-levels of expression, and because maximal degeneracy would result in 5^{10} nucleotide sequences for the Mam2 i3 loop, the codons targeted were not made maximally degenerate. This allowed more codons to be targeted while keeping the degeneracy experimentally tractable. DNA transformation

bR i3

1) **ser lys ala** ..... **glu ser met arg pro**
2) T*cg aag g*CC ..... GAA A*gc atg c*GC CCC

bR:Mam2 i3

3) **ser lys ala arg ala ile leu ile arg lys lys ile gly gln phe pro phe gly pro met gln cys ile leu ser met arg pro**
4) T*cg aag g*CC cgc **gcA** atc ctg atc cga **aaG aaA atT ggC caG ttC ccA ttT ggT ccA** ATG **caG tgT atC ctC** A*gc atg c*GC CCC
G **A G C T A T G C C G A C T T**

					Sequences
5)	2 x	2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2 x 2	x	2 x 2 x 2 x 2	= 3.3×10^4
				2^{15}	= 3.3×10^4

Figure 7: Design of bR:Mam2 DCM oligonucleotides

The amino acids are shown purple in line 1, and the DNA coding sequences are shown black in line 2, for the i3 loop of bacteriorhodopsin. The upstream StyI and downstream SphI cloning sites are shown underlined and italicized in line 2. The amino acids in line 3 (bR:purple and Mam2:blue) and DNA sequences in line 4 (black) were used to for the oligonucleotide-directed mutagenic construction of the *bop:mam2* i3 loop fusion genes. Codons targeted for degenerate mutagenesis are shown in bold with the wobble bases used shown in red. The upstream StyI and downstream SphI cloning sites are shown underlined and italicized in line 4. The total number of bR:Mam2 i3 loop coding sequences in the degenerate codon mutagenesis strategy is 3.3×10^4, as calculated in line 5.

Oligo #	Sequence (5′ to 3′)
1	CA T**CC AAG G**CC CGC GCR ATC CTG ATC CGA AAR AAR ATY GGGY CAR TTY CCR TTY GGY CCR ATG CAR TGY ATY CTY A**GC ATG C**AA CGG
2	CCG TTG CAT GCT RAG RAT RCA YTG CAT YGG RCC RA YGG TAA YTG RCC RAT YTT YTT TCG GAT CAG GAT YGC GCG GGC CTT GGA GT
3	GTG CTG TTC TTC GGG TTC ACC T***C*** **AAG G**CC GAA A**GC ATG C**GC CCC GAG GTC GCA TCC
4	GGA TGC GAC CTC GGG GC**G CAT GC**T TTC GG**C CTT *GG***A GGT GAA CCC GAA GAA CAG CAC
5	CCG TTC GGT GGG GAG CAG AAC C
6	CGA TCA GCC ACA CGA CGG GA
7	CCCAT T***C*** **AAG G**CG G***CT TAA G***AT <u>TGA</u> **GC ATG C**CA AGCGC
8	GCGCA TG**G CAT GC** <u>TCA</u> AT**C TTA AG**C CG**C CTT GG**A ATGGG

Table I. Oligonucleotides Used in Degenerate Codon Mutagenesis

The oligonucleotides (oligos) used in our degenerate codon mutagenesis are numbered from 1-8 and listed 5′ to 3′. The coding and noncoding strands of the DCM Mam2 i3 loop region are shown as oligo #1 and #2, respectively. Oligo 1 shows the StyI (CCAAGG) and SphI (GCATGC) restriction sites in bold. Oligos #3 (coding) and #4 (noncoding) show a silent Sty1 restriction site, in bold, that was engineered into the pENDs bop 8 vector in order to facilitate ligation with oligos 1 and 2. Oligos #5 and #6 were used to sequence the individual DNA of all *E. coli* transformants that resulted from the ligation of oligos 1 and 2 with 3 and 4. Oligos #7 and #8 show the construction of stuffer fragments used to eliminate the bR^+ background caused by pENDs bop 51. Oligo #7 shows the StyI (CCAAGG), AflII (CTTAAG), and SphI (GCATGC) restriction sites in bold and the in frame stop codon (TGA) underlined.

1st Position	2nd Position				3rd Position
	U	C	A	G	
U	0.18 Phe	0.11 Ser	0.24 Tyr	0.31 Cys	U
	2.92 Phe	1.20 Ser	2.28 Tyr	0.42 Cys	C
	0.07 Leu	0.22 Ser	0.07 STOP	0.18 STOP	A
	0.66 Leu	1.86 Ser	0.10 STOP	1.07 Trp	G
C	0.33 Leu	0.10 Pro	0.19 His	0.40 Arg	U
	3.97 Leu	1.81 Pro	2.05 His	3.31 Arg	C
	0.11 Leu	0.35 Pro	0.37 Gln	0.46 Arg	A
	3.19 Leu	2.42 Pro	2.23 Gln	2.11 Arg	G
A	0.21 Ile	0.27 Thr	0.11 Asn	0.25 Ser	U
	3.28 Ile	3.18 Thr	2.01 Asn	1.59 Ser	C
	0.07 Ile	0.38 Thr	0.57 Lys	0.08 Arg	A
	1.74 Met	2.98 Thr	1.08 Lys	0.12 Arg	G
G	0.64 Val	0.51 Ala	0.90 Asp	0.73 Gly	U
	5.00 Val	5.73 Ala	8.09 Asp	4.98 Gly	C
	0.18 Val	0.82 Ala	1.57 Glu	0.48 Gly	A
	3.76 Val	6.03 Ala	5.11 Glu	2.30 Gly	G

Table II. Codon Usage in *H. salinarum*

This table represents the varying percentages of codon usage in *H. salinarum*. Codons used in our degenerate codon mutagenesis (DCM) are highlighted in yellow and rare codons not used in our DCM are red. The sequence of all sixty-four codons will lead you to the corresponding amino acid (three letter abbreviation) and its representative percentage in the genome of *H. salinarum*. Simply take the first nucleotide of the codon and match it with the correct row under the column labeled **1st Position**. Once in the correct row, move across the table until you are in the correct **2nd Position** column of the second nucleotide in the codon. Finally, match the third nuceleotide of the codon with the row of the corresponding nucleotide in the **3rd Position** column. You should now have identified the amino acid corresponding to your codon and its corresponding percentage used in the *H. salinarum* genome.

This table was adapted from the Center for Biological Sequence Analysis at the Technical University of Denmark (DTU) website [**88**].

efficiencies for *H. salinarum* strain L33 range between 10^4-10^6 transformants per microgram of supercoiled plasmid DNA. Therefore, a library of approximately 10^4 DCM sequences was designed. In cases of amino acids with more than two synonymous codons in the genetic code, we excluded the least abundant codon but included the lesser abundant codon to avoid generating extremely stable secondary structures that can originate from a very high content G:C biased codons. Using stable secondary structures would make subsequent molecular engineering manipulations problematic (eg. annealing, cloning, sequencing). In addition, previous work did not indicate direct correlation between GC wobble base bias and high-levels of gene expression. The degenerate oligos for the coding and non-coding strands were purchased from Oligomer, OY, Helsinki, FI.

Oligo Annealing

All nucleic acid concentrations were determined by optical density at 260 nanometers. The oligos were diluted to 3 microgram per microliter in deionized water. 10 micrograms of each oligo was mixed in 400 microliters (total volume) of 0.5% phosphate buffered saline and 0.5 millimolar EDTA at pH 7. The solution was heated to 95°C for 30 minutes in an aluminum heat block and slow cooled to 30°C, reheated to 70°C for 15 minutes and slow cooled to room temperature. The annealing solution was then digested with exonuclease I (NEB, Ipswich, MA) to remove annealed and single-stranded oligomers. The annealed oligos were purified on NucleoSpin Extract II columns (Macherey-Nagel, Gutenburg, FR) and digested with StyI and SphI. The restriction enzymes were heat inactivated at 60°C for 20 minutes and slow cooled to room temperature. Annealed

oligonucleotide concentration was determined from absorbance at 260 nanometers using a BioPhotometer (Eppindorf, Westbury, NY).

Cloning the Mam2 Degenerate Codon Mutagenesis i3 loop

Oligo-directed mutagenesis was used to engineer a silent StyI DNAse restriction site into the pENDS bop 8 vector [**77**]. The site was introduced after amino acid position 163 in the *bop* gene i3 loop coding-region (**Table I: 3, 4**). DNA from a positive transformant was isolated and the construction confirmed by digestion with StyI. The resulting vector was called pENDS bop 51. pENDS bop 51 was digested preparatively with StyI and SphI. The resultant 3.9 Kb fragment was gel-purified and extracted from agarose with NucleoSpin Extract II columns (Macherey-Nagel, Gutenburg, FR). The StyI and SphI digested Mam2 i3 loop DCM oligos and purified vector fragments were ligated and transformed into *E.coli* strain DH5α. Colony counting indicated a yield of ~5 x 10^5 transformants, covering the designed degeneracy. 175 individual transformants were isolated and plasmid DNA was purified using the Promega Wizard® Plus SV Miniprep Kit. The individual DNA's were sequenced in both orientations (**Table I: 5, 6**) to evaluate the DCM sequences (**Table III**, DNA Sequencing Facility, BioCenter, University of Helsinki, Helsinki, FI).

DNA sequencing indicated that 50% of the transformants possessed unique in-frame Mam2 DCM i3 loop coding regions. 40% of the remaining transformants contained missensence mutations, which will lead to null expression phenotypes. The remaining 10% of the transformants possessed the *bop* gene i3 loop coding region. Since bR is highly expressed in *H.*

1) **ser lys ala** ..... **glu ser met arg pro**

2) cg aag gcc ..... gaa agc atg cGC CCC
3) *cc* aag gcc ..... **gaa** agc atg cGC CCC

4) **ser lys ala arg ala ile leu ile arg lys lys ile gly gln phe pro phe gly pro met gln cys ile leu ser met arg**
S K A R A I L I R K K I G Q F P F G P M Q C I L S M R

06 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaG atC ggC caG ttC ccA ttT ggT ccA** ATG **caG tgT atC ctC** *agc atg c*
07 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaA atT ggC caA ttT ccA ttT ggC ccA** ATG **caA tgT atT ctC** *agc atg c*
08 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaA atT ggC caG ttT ccG ttC ggG ccA** ATG **caA tgC atC ctC** *agc atg c*
09 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaG aaG atT ggT caA ttT ccA ttC ggC ccA** ATG **caG tgC atT ctC** *agc atg c*
12 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaA atC ggT caG ttT ccA ttT ggT ccA** ATG **caA tgC atT ctT** *agc atg c*
14 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaG atT ggC caA ttC ccG ttC ggC ccG** ATG **caG tgC atT ctC** *agc atg c*
17 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaG atT ggT caA ttT ccG ttC ggC ccA** ATG **caG tgC atT ctC** *agc atg c*
19 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaG atT ggT caG ttC ccA ttC ggT ccA** ATG **caG tgC atC ctC** *agc atg c*
21 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaG atC ggC caA ttT ccA ttT ggC ccG** ATG **caA tgT atT ctT** *agc atg c*
22 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaA atT ggC caG ttC ccA ttC ggT ccG** ATG **caA tgC atC ctC** *agc atg c*
23 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaA atT ggC caG ttT ccA ttT ggT ccA** ATG **caG tgC atT ctC** *agc atg c*
24 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaA atT ggC caG ttC ccA ttC ggT ccG** ATG **caA tgC atC ctC** *agc atg c*
27 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaG aaG atC ggT caA ttC ccA ttT ggC ccG** ATG **caA tgC atT ctT** *agc atg c*
28 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaG atT ggT caA ttT ccA ttT ggT ccG** ATG **caA tgT atT ctT** *agc atg c*
29 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaA aaA atT ggT caA ttT ccG ttC ggT ccG** ATG **caA tgC atC ctT** *agc atg c*
31 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaG atT ggC caG ttT ccA ttC ggC ccA** ATG **caA tgC atT ctC** *agc atg c*
36 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaA atC ggC caA ttT ccA ttT ggC ccG** ATG **caG tgT atT ctC** *agc atg c*
38 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaA atC ggC caA ttT ccG ttT ggC ccA** ATG **caA tgC atC ctT** *agc atg c*
40 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaG atC ggT caA ttC ccG ttC ggT ccA** ATG **caG tgC atT ctC** *agc atg c*
41 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaG atT ggC caG ttT ccG ttC ggC ccA** ATG **caA tgC atC ctC** *agc atg c*
43 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaA aaG atT ggC caG ttC ccA ttT ggT ccA** ATG **caG tgT atC ctT** *agc atg c*
45 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaG atC ggT caG ttT ccA ttT ggT ccA** ATG **caG tgC atT ctC** *agc atg c*
46 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaG atT ggT caA ttC ccG ttT ggC ccG** ATG **caA tgT atC ctT** *agc atg c*
47 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaA atC ggT caA ttC ccA ttT ggC ccG** ATG **caG tgT atC ctC** *agc atg c*
49 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaA atC ggT caG ttC ccA ttC ggT ccG** ATG **caA tgT atC ctC** *agc atg c*
55 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaG atC ggT caA ttC ccG ttC ggC ccA** ATG **caA tgC atT ctC** *agc atg c*
57 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaA aaG atT ggC caA ttC ccG ttT ggT ccG** ATG **caA tgC atT ctC** *agc atg c*
59 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaG atC ggT caA ttC ccA ttC ggT ccA** ATG **caG tgC atC ctC** *agc atg c*
62 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaA aaA atC ggC caG ttC ccG ttC ggT ccA** ATG **caA tgT atC ctT** *agc atg c*
64 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaA aaA atC ggC caA ttC ccG ttC ggC ccG** ATG **caA tgT atC ctT** *agc atg c*
65 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaA atT ggC caG ttT ccG ttC ggT ccA** ATG **caA tgT atT ctC** *agc atg c*
67 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaA atT ggT caA ttC ccA ttT ggT ccA** ATG **caG tgC atC ctT** *agc atg c*
68 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaA aaA atC ggC caG ttC ccA ttC ggC ccA** ATG **caA tgT atT ctC** *agc atg c*
69 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaG aaA atT ggC caG ttC ccA ttT ggT ccG** ATG **caG tgT atT ctT** *agc atg c*
70 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaG atT ggC caG ttT ccA ttC ggC ccA** ATG **caA tgC atT ctC** *agc atg c*
72 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaG atC ggT caG ttT ccA ttT ggT ccA** ATG **caG tgC atT ctC** *agc atg c*
76 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaA aaA atC ggT caG ttT ccG ttT ggC ccG** ATG **caG tgC atC ctT** *agc atg c*

83 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaG aaG atC ggC caG ttT ccA ttT ggC ccG** ATG **caA tgT atC ctT** *agc atg c*
85 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaA atC ggC caG ttC ccG ttC ggC ccG** ATG **caA tgT atT ctC** *agc atg c*
87 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaG aaA atC ggT caG ttC ccA ttC ggC ccG** ATG **caA tgC atC ctC** *agc atg c*
91 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaG aaA atT ggC caA ttC ccA ttC ggC ccA** ATG **caA tgT atT ctC** *agc atg c*
93 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaA atT ggT caG ttT ccA ttC ggC ccG** ATG **caA tgC atC ctC** *agc atg c*
96 *cc aag* gcc cgc **gcA** atc ctg atc cga **aaG aaA atC ggT caG ttT ccA ttT ggC ccG** ATG **caG tgT atT ctT** *agc atg c*
98 *cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaG atC ggC caG ttT ccG ttC ggC ccG** ATG **caG tgC atT ctT** *agc atg c*
100*cc aag* gcc cgc **gcA** atc ctg atc cga **aaG aaG atT ggT caG ttC ccA ttT ggT ccA** ATG **caG tgT atC ctC** *agc atg c*
101*cc aag* gcc cgc **gcA** atc ctg atc cga **aaA aaG atT ggC caG ttT ccG ttC ggC ccA** ATG **caG tgC atC ctC** *agc atg c*
104*cc aag* gcc cgc **gcA** atc ctg atc cga **aaG aaA atC ggT caG ttT ccA ttT ggC ccG** ATG **caG tgT atT ctT** *agc atg c*
105*cc aag* gcc cgc **gcA** atc ctg atc cga **aaG aaG atC ggC caG ttT ccA ttT ggC ccG** ATG **caA tgT atC ctT** *agc atg c*
121*cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaG atC ggC caG ttT ccA ttT ggC ccG** ATG **caG tgT atT ctT** *agc atg c*
122*cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaA atT ggC caG ttT ccA ttC ggC ccG** ATG **caA tgC atT ctC** *agc atg c*
123*cc aag* gcc cgc **gcA** atc ctg atc cga **aaA aaA atT ggC caA ttC ccA ttT ggT ccG** ATG **caG tgC atT ctC** *agc atg c*
124*cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaG atC ggT caA ttT ccG ttT ggC ccG** ATG **caG tgC atC ctT** *agc atg c*
126*cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaA atT ggT caA ttT ccG ttT ggT ccG** ATG **caG tgT atT ctT** *agc atg c*
127*cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaG atT ggC caA ttC ccG ttC ggT ccA** ATG **caA tgC atT ctC** *agc atg c*
128*cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaA atT ggC caG ttC ccG ttC ggT ccG** ATG **caG tgC atT ctC** *agc atg c*
129*cc aag* gcc cgc **gcA** atc ctg atc cga **aaA aaA atC ggC caG ttC ccA ttT ggC ccG** ATG **caG tgT atT ctC** *agc atg c*
130*cc aag* gcc cgc **gcG** atc ctg atc cga **aaG aaA atT ggC caA ttC ccG ttT ggT ccG** ATG **caA tgT atT ctC** *agc atg c*
133*cc aag* gcc cgc **gcA** atc ctg atc cga **aaA aaG atT ggT caA ttT ccG ttC ggC ccA** ATG **caG tgT atT ctC** *agc atg c*
136*cc aag* gcc cgc **gcA** atc ctg atc cga **aaA aaA atC ggC caG ttC ccA ttC ggT ccA** ATG **caA tgT atC ctT** *agc atg c*
138*cc aag* gcc cgc **gcG** atc ctg atc cga **aaA aaG atT ggT caG ttC ccG ttC ggT ccA** ATG **caA tgC atC ctT** *agc atg c*

Table III. All Degenerate Sequences of *bop:mam2* Fusions

The degenerate sequences of our *bop:mam2* i3 loop fusions are shown. Line 1: The 8 amino acids of the bR i3 loop are shown in purple. Line 2: The i3 loop coding region of bR. Line 3: The i3 loop coding region of bR showing the silent mutation of the engineered SphI restriction site, in red. Line 4: Amino acid sequence of our bR:Mam2 i3 loop fusion protein, with bR amino acids in purple, Mam2 amino acids in blue, and one letter abbreviations directly below. All degenerate sequences, numbered from 06 to 138, having degenerate codons in bold and their corresponding wobble bases in red.

salinarum a background of 10% *bop* gene i3 loop coding region will lead to bR^+ expression phenotypes (intensely purple colonies) and complicate identification of highly expressed DCM bR:Mam2 fusion receptors. Therefore, the DCM cloning strategy was modified. A stuffer fragment was designed that included upstream *Sty*I and downstream *Sph*I cloning sites, a unique and internal DNA restriction site (*Afl*II), and an in-frame translation stop codon (**Table I, oligos 7 and 8**). The stuffer fragment oligos were annealed and purified as per above. The *Sty*I and *Sph*I digested stuffer and pENDS bop 51 vector fragments were ligated and transformed into *E.coli* strain DH5α. DNA from a positive transformant was isolated and the construction confirmed by digestion with *Afl*II and DNA sequencing. The resulting vector was named pENDS bop 52. pENDS bop 52 was digested with *Sty*I and *Sph*I and the 3.9 Kb vector fragment gel purified. The *Sty*I and *Sph*I digested Mam2 i3 loop DCM oligos and purified vector fragments were ligated and transformed into *E.coli* strain DH5α. Colony counting again indicated ~5 x 10^5 transformants, covering the designed degeneracy. DNA was isolated and sequenced from 50 colony purified transformants. Again, 50% of the transformants possessed unique Mam2 DCM i3 loop coding regions and 10% of the transformants possessed the stuffer fragment. Due to the in-frame translation stop codon, this background would produce bR^- expression phenotypes (white colonies). The remaining transformants were missense mutations and would also lead to bR^- expression phenotypes.

Expression vector subcloning

H. salinarium expression vectors (pHex) were based on the *E. coli*/haloarchaeal shuttle vector pUBP2 [**63,66,77,87**]. pHex

vectors possess replicons derived from *E. coli* and *H. salinarium*, and genes which confer resistance to selectable markers in both hosts, ampicillin (transpeptidase inhibitor) and mevinolin (HMG-CoA reductase inhibitor), respectively. 4 micrograms of each of 34 individual pENDS DCM clones were digested with *Pst*I and *Bam*HI (**Figure 8**). The resulting 1.2 Kb fragment containing the *bop:mam2* gene fusion was gel purified and subcloned into the P*st*I and B*am*HI digested pHex. Positive clones were confirmed by restriction digestion and 10 were sequenced to confirm the anticipated gene fusion. Pooled DCM DNA was also isolated from 5 ml liquid cultures of pENDS bop 52 DCM i3 loop transformants. The pooled DNA was digested with *Pst*I, *Bam*HI, and A*fl*II and the desired 1.2 Kb *bop:mam2* gene fusion construction gel purified. Low intensity staining of the smaller P*st*I-A*fl*II and A*fl*II-B*am*HI DNA fragments confirmed a minor background of stuffer fragment clones. Simultaneous digestion with A*fl*II served to further reduce this background. Three aliquots of 1.5 micrograms of each pHex DNA (individual or pool) was dried in a Savant Speed-Vac (Holbrook, NY) and resuspended in 15 uL of deionized water for transformation into *H. salinarum*. The desired clones were confirmed by restriction digestion with P*st*I and B*am*HI (**Figure 8**).

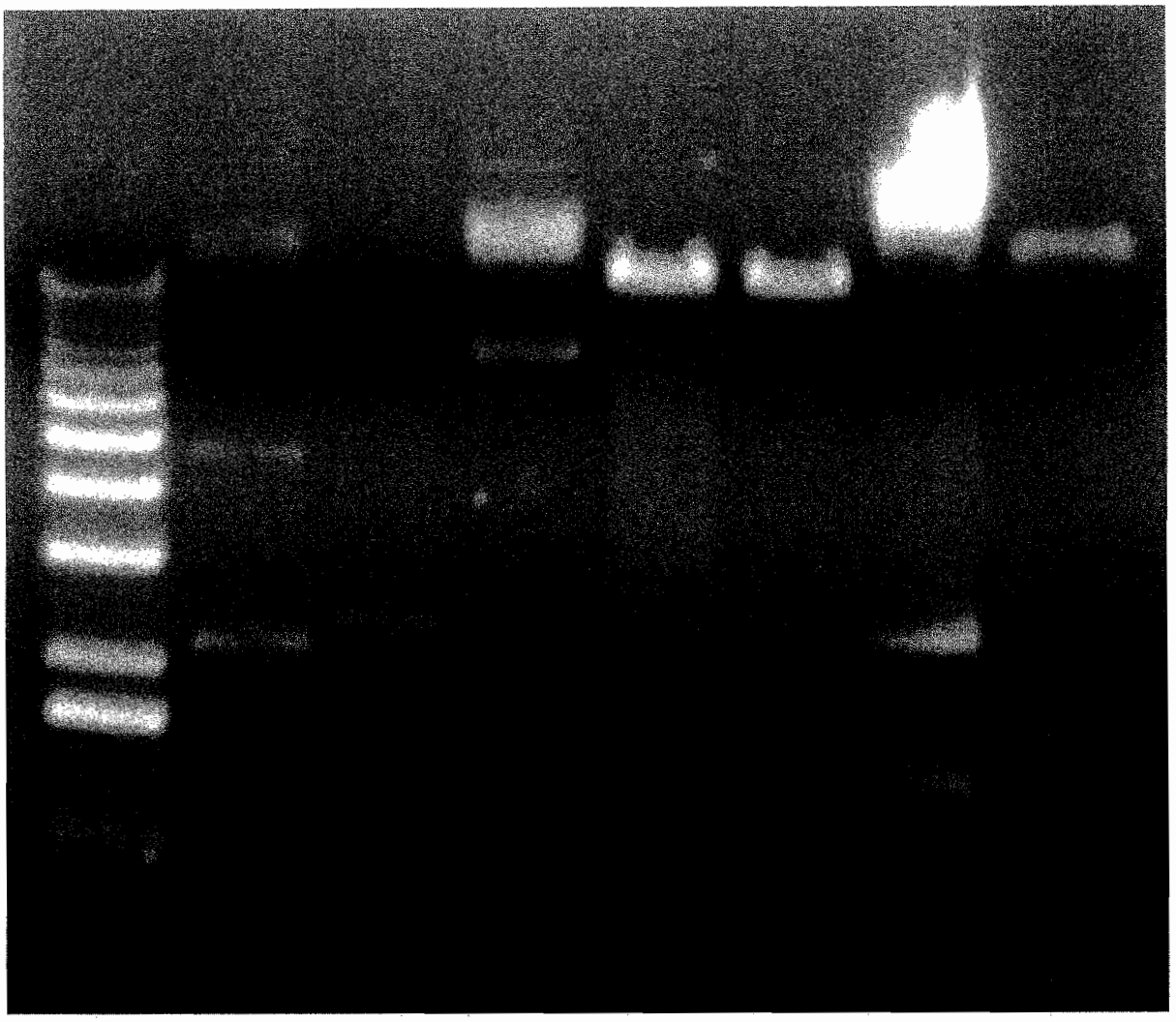


Figure 8. Gel Purification of Restriction Digests

Representative results of all gel purified restriction endonuclease digests. Lane 1: molecular markers. Lane 2: pENDS vector uncut. Lane 3: pENDS vector digested with PstI and BamHI restriction endonucleases. Lane 4: pHex shuttle vector uncut. Lane 5: pUBP2 vector resulting from restriction endonuclease digest of pHex shuttle vector. Lane 6: PstI and BamHI restriction endonuclease digest of pHex vector after ligation of 1.2 kilobase *mam2* i3 insert. Lane 7: pHex pool plasmid containing a pENDS-type vector contaminant. Lane 8: pHex pool plasmid after purification.

RESULTS AND DISCUSSION

The bR Scaffold

The ultimate goal of this work is to clone a bR:GPCR gene fusion that can facilitate robust production of chimeric protein in large enough quantity for structural and functional studies. How do we go about obtaining such quantities of quality protein to ensure successful structural and functional characterization? We chose to insert our GPCR domain into a protein scaffold well-characterized and well-suited for structural study: bacteriorhodopsin (bR). bR is structurally homologous to GPCR containing 7 transmembrane spanning domains, is highly expressed in the archaeon *Halobacterium salinarum* (*H. salinarum*), and forms highly resolvable crystals, providing an ideal scaffold for obtaining structural information at the atomic level of bR containing fused GPCR domains.

Assignment of Mam2 i3 Loop Sequence

In order to design our chimeric bR:Mam2 constructs, we sought to assign membrane topography based on current literature and existing GPCR structures. This receptor in *S. pombe* was chosen due to the proven utility of yeast systems in the study of GPCR [**49-52**], the association of only a Gα monomeric signaling protein [**55**], and the existence of efficient *in vivo* signaling assays [**56**]. As discussed earlier, we putatively assigned TM domains and extramembranous regions of the Mam2 receptor, and identified specific amino acids included in the i3 loop of Mam2 (**Figure 5, 6**). The proposed Mam2 secondary structure shows 7-transmembrane spanning domains similar to

that of bR, but having longer N- and C-termini. More importantly, our putative structural assignment provided identification of the amino acids comprising the i3 loop of Mam2. This information allowed for custom *mam2* i3 oligonucleotides to be constructed and inserted into the *bop* gene, which can now be cloned and expressed as chimeric bR:Mam2 proteins.

The Third Intracellular Loop

Mutagenesis has indicated the i2 through i4, and to a lesser extent the i1, of GPCR as critical for signal transduction; however the specific role of each varies from receptor to receptor. In this work we chose to manipulate the i3 loop of bR by replacing it with the corresponding i3 loop of Mam2. Molecular engineering will be facilitated by the relatively small size of the Mam2 i3 loop. Although there is no evidence to link the i3 loop of Mam2 to signal transduction in *S. pombe*, previous mutagenesis studies and comparison with other pheromone receptors have implicated the i3 loop as a critical component in signal transduction.

i3 Loop Rationale

The i3 loop is responsible for signaling across various GPCR types, such as the 5-HT_6 serotonin receptor and the α2-adrenergic receptor [**42,43**]. Truncation experiments reveal that signal transduction in the homologous Ste2p pheromone receptor of *S. cerevisiae* is <u>not</u> inhibited by complete or partial removal of the C-terminus [**47**]. Additionally, the observation of conserved i3 loop length in pheromone receptors, as well as limited mutagenesis of amino acid residues therein, indicates a key role for the Mam2 i3 loop

interactions with Gpa1 during signal transduction in *S. pombe* [**57**]. This forms the basis of our hypothesis that the i3 loop is critical for Gpa1 coupling in Mam2 and will be used to test our bR:Mam2 fusion receptor strategy.

Successful expression of our bR:Mam2 chimeras in *H. salinarum* is feasible considering successful expression experiments of other fusion proteins in the same system. For example, a fusion protein of the bR scaffold containing the catalytic subunit of the Aspartate Transcarbamylase protein of *E. coli* fused to the C-terminus of bR was successfully expressed in *H. salinarum*. The colonies formed were purple indicating proper folding into the cell membrane, and isolation by sucrose density centrifugation afforded about 3 milligrams per liter of cell culture [**50**]. Another experiment successfully expressed a chimera containing the i3 loop region of the rat $\alpha 2_B$-adrenergic receptor fused into the bR scaffold, replacing the i3 loop of bR. Again, expressed in *H. salinarum*, 0.5 milligrams of purple protein per liter cell culture was isolated and spectroscopic analysis revealed a properly folded and fully functional light-driven proton pump [**80**].

Functional expression of chimeric proteins containing the bR scaffold provides even further utility. The expression of a bR:$\alpha 2_B$-adrenergic fused, fully functional photoreceptor in *H. salinarum* has been described [**80**]. Additionally, various loop regions of bR were replaced with the 13 amino acid C-terminus of the Sendai virus L-protein and expressed in either *E. coli* or *S. pombe* [**44**]. Functional photoreceptors were expressed in four separate fusions containing the non-native peptide at the e1, e2, i2 and i3 positions.

In addition to expression of bR chimeras in *H. salinarum*, and functional expression of proton-pumping bR chimeras in *H. salinarum* and *S. pombe*, bR chimeras can induce signal transduction through light activation [**23**]. The i3 loop of bR was replaced with that of the GPCR bovine rhodopsin (Rho), and expressed in *H. salinarum*. The resulting chimera, aside from being purple and properly folded into the membrane, was able to increase the basal activation of the transducin Gα subunit in *H. salinarum*. This potential for light-activated signal transduction in bR:GPCR chimeras gives promise to our goal of studying the intricacies of signal transduction in *S. pombe* using our bR:Mam2 chimeras.

The Mam2 Secondary Structure and i3 Loop Sequence

The Mam2 secondary structure shows 7TM domains characteristic of all GPCR (**Fig. 6**). The N-terminus, C-terminus, extracellular (e1, e2, and e3), and intracellular (i1, i2, and i3) loops of Mam2 are significantly longer than bR (**Fig. 4**). This is attributed to properties required for signal transduction in GPCR, ligand binding site(s) and G-protein interacting domains. The Mam2 i3 loop is 24 amino acids in length, three times longer than the i3 loop of bR (**Figure 9**) and contains polar and non-polar residues to facilitate selective coupling to Gpa1. The N-terminal end of the Mam2 i3 loop, proximal to TM5, contains a lysine (K) and a leucine (L) which is similar to the lysine and alanine (A) of bR. The Mam2 i3 loop also contains a cysteine (C) that will be useful in future G-protein interaction studies, discussed previously.

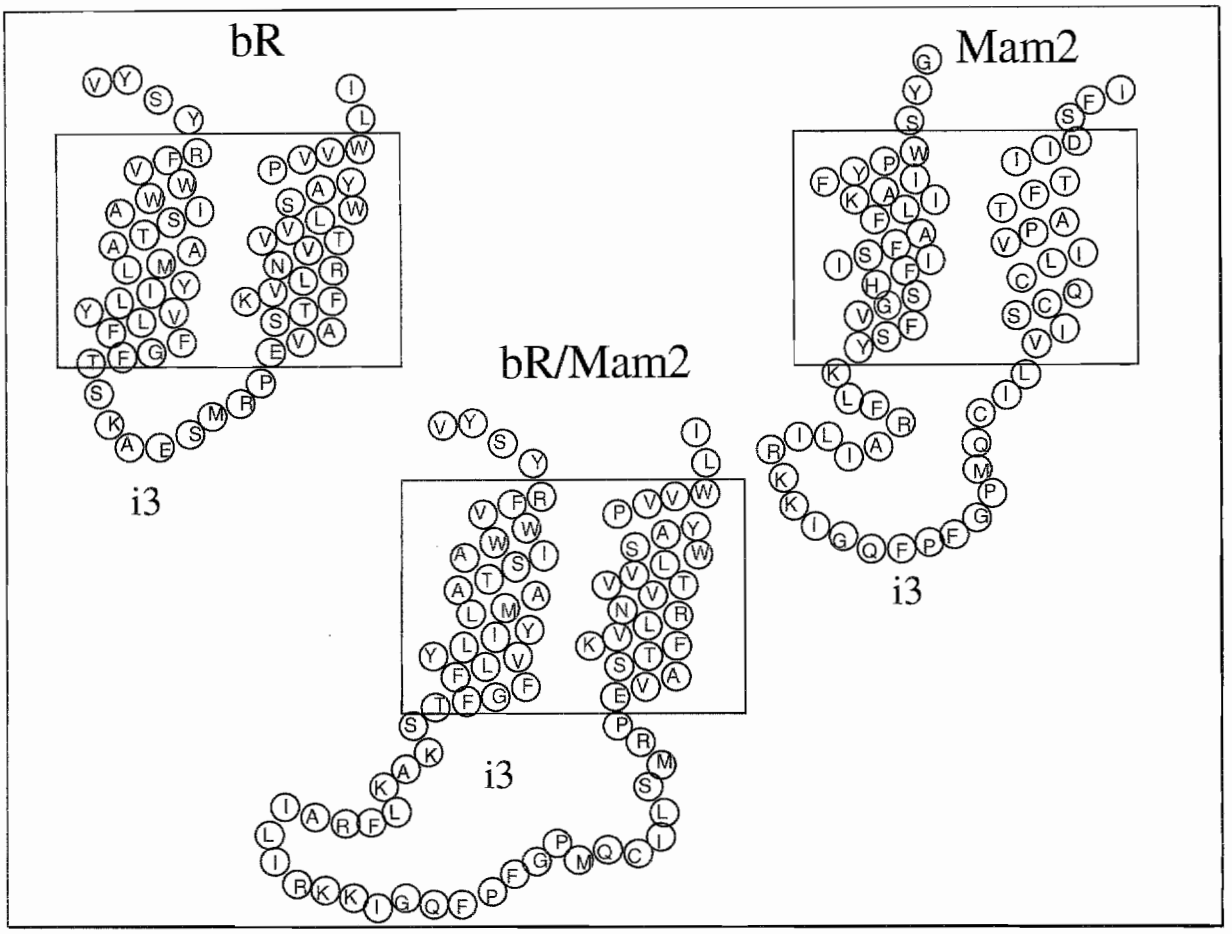


Figure 9: The bR:Mam2 Fusion Strategy

Our bR:Mam2 fusion proteins were developed using this general scheme. The small black boxes indicate the cell membrane and amino acids are denoted as their corresponding one letter abbreviation. The middle image represents the bR:Mam2 i3 loop fusion. The bR TM5 and TM6 are shown in green and the Mam2 i3 loop is shown in blue. The upper left shows the bR i3 loop, in red, connecting the bR TM5 and TM6, shown in green. The upper right shows the Mam2 i3 loop, in blue, connecting the Mam2 TM5 and TM6, shown in black.

bR:Mam2 Fusion Strategy

The bR:Mam2 i3 loop fusion receptor strategy was selected to facilitate the biochemical and biophysical analysis of GPCR structure and function. The bR:Mam2 fusion receptor will be made by replacing the i3 loop of bR with that of Mam2 (**Figure 9**). To construct this fusion, the i3 loop nucleotide sequence of *mam2* was inserted into the *bop* gene by engineering a silent mutation at position 163 of the *bop* gene i3 loop coding region, introducing a StyI restriction endonuclease site. Together with the SphI restriction endonuclease site at the 3' end of the *bop* gene i3 loop, the *mam2* could be inserted or "ligated" into the gene (**Fig. 7**). Therefore the first three amino acids (serine, ser, S; lysine, lys, K; and alanine, ala, A) and the last four amino acids (serine, ser, S; methionine, met, M, arginine, arg, R; and praline, pro, P) of bR were included in our bR:Mam2 i3 fusion. The first two amino acids of the Mam2 i3 loop, lysine (lys, K) and leucine (leu, L) were omitted due to similarities with the corresponding lysine (lys, K) and alanine (ala, A) of bR (**Fig. 9**). Finally, the third amino acid of the Mam2 i3 loop, phenylalanine (phe, F) was removed to facilitate ligation with *bop* gene. All Mam2 i3 amino acids at the C-terminal end, proximal to TM6, were included. The sequences of all degenerate *bop:mam2* fusions are shown (**Table III**).

The fused *bop:mam2* gene sequences were transformed into *E. coli* and verified by DNA sequencing. 50% of the transformants successfully possessed our unique *mam2* i3 coding regions, while 40% contained missense mutations that will lead to null expression phenotypes, and 10% contained *bop* i3 coding region. This 10% background will express bR^{+} phenotypes complicating

identification of highly expressed bR:Mam2 fusion receptors in *H. salinarum*, and results from gel purification lacking the ability to resolve uncut pENDS DNA from digested pENDS DNA. To eliminate this bR^+ background, a stuffer fragment was designed and inserted into pENDS containing a unique DNA restriction site, AflII, and an in-frame translation stop codon. The background of this new vector, pENDS bop 52, will result in a bR^- expression phenotype due to the in-frame translation stop codon. Again, the fused *bop:mam2* gene sequences were transformed into *E. coli* and verified by DNA sequencing. 50% of transformants in *E. coli* possessed our unique *mam2* i3 coding regions, while 40% contained missense mutations (bR^-), and 10% contained pENDS bop 52 (bR^-). Successful transformants were cloned into the pHex shuttle vector and transformed into *E. coli* to obtain suitable DNA concentrations for transformation into *H. salinarum* (**Table IV**). This represents the essential step toward preparing *bop:mam2* i3 loop DNA for transformation into *H. salinarum* and expression of our bR:Mam2 fusion protein therein.

Degenerate Codon Mutagenesis of DNA Sequences

In our system, 34 degenerate DNA sequences of Mam2 i3 loops were cloned into the *bop* gene in the shuttle vector pHex in preparation for transformation in *H. salinarum* (**Table IV**). The presence of purple colonies would indicate which sequences are being expressed, providing insights as to the factors involved in protein expression.

How is bR expressed in such large quanities in *H. salinarum*? Despite how well understood bR is from a structural and functional standpoint, the specific factors that promote

Clone #	i3 pENDS (ng/µl)	i3 pHex (ng/µl)
6	162	320
7	165	160
8	166	280
9		
12	185	
14	202	
17	96	TBD
18	104	238
29	142	
31	95	
35	118	
36	83	8
38	116	46
41	92	368
42	97	TBD
43	124	TBD
44		
45	95	TBD
46	169	TBD
47	93	TBD
49	110	TBD
55	173	629
57	103	476
58	178	927
62	134	1098
63	113	687
66	111	262
67	25	70
68	111	
69	185	400
70	111	700
72	117	286
76	102	726
83	116	
85	125	
87	113	
91	122	
93	102	270
96	116	
98	97	10

Clone #	i3 pENDS (ng/µl)	i3 pHex (ng/µl)
100	87	
101	156	175
104	85	
105	69	
109		
114		
115		
120		
122		
123	131	
124	116	
125		
127		
128	116	75
129	130	240
130	139	
133	95	
136	94	
138	94	
140	-------	
141	-------	
145	-------	
152	96	
153	120	
154	105	
155	101	
156	112	
157	96	
158	89	
159	107	
160	100	
161	114	
162	112	
163	118	
164	104	
165	121	
166	116	
167	118	
168	112	

Table IV. Cloning of Degenerate *bop:mam2* Sequences
The 34 sequences successfully cloned into the pHex shuttle vector are shown in light blue, with concentrations of i3 pENDS DNA prior to cloning shown in the second column. The resulting concentrations of pHex plasmids containing the degenerate *bop:mam2* i3 sequence ready for transformation into *H. salinarum* are shown in the third column. Concentrations labeled "TBD" will be determined.

expression at such high levels remains unknown. Critical non-coding regions in the *bop* gene may promote large accumulation of mRNA, regulate the organized folding of protein into the cell membrane, or stabilize the mRNA tertiary structures formed. However, without this information, degenerate codon mutagenesis serves as a tool to express multiple degenerate genes and allow the expression system to phenotypically indicate preferred sequences.

Why clone 34 degenerate mutants instead of simply cloning an entire pool to obtain expression? The cloning of degenerate *bop:mam2* gene fusions will allow for potential high expression levels of our fusion protein as well as provide an indication as to what genetic factors contribute to high expression levels. Successful expression of bR:GPCR fusions has been accomplished, as discussed previously. However, factors contributing to, or controlling the level of expression have yet to be determined. Using degenerate *bop:mam2* gene fusions we will be able to identify genetic differences between transformants of *H. salinarum* expressing varying levels of our bR:Mam2 fusion protein, ultimately leading to insights into the genetic factors regulating expression.

As mentioned earlier, high levels of mRNA accumulation and efficient translation are likely caused by information contained in the coding region of wild-type *bop* mRNA. Since our chimeric receptor strategy involves replacing the *bop* gene coding region with coding regions of non-native *mam2* genes, critical sequence determinants contributing to *bop* gene expression may be perturbed. The use of degenerate genes allows for identification of perturbations that will affect expression while not affecting the amino acid sequence of our

protein. We have cloned 34 degenerate mutant coding sequences into the *bop* gene in an attempt to express chimeras containing minimal perturbations to the regions regulating gene expression. This work represents the essential first step toward expression of chimeric bR:Mam2 proteins in *H. salinarum*.

The library of DCM sequences was based on covering the codon bias in *H. salinarum* by including both high and low frequency codons. Codons for alanine (Ala), isoleucine (Ile), glycine (Gly), proline (Pro), and leucine (Leu) were chosen by taking the highest frequency and the second lowest frequency codons, excluding rare and relative intermediate frequency codons, as in the case with ATA for Ile (0.07%) and CTG for Leu (3.19%), respectively **(Table II)**. Both degenerate codons for amino acids such as phenylalanine (Phe), glutamine (Gln), lysine (Lys), and cysteine (Cys) were used, as each amino acid only corresponds to two codons. The methionine (Met) residue contains no degeneracy, corresponding to only a single codon.

Why not use only high frequency codons? It is logical to think that using only high frequency codons would result in high levels of expression. However, this is not so for the rat angiotensin type-1A (AT1A) GPCR, or for bR. Both contain a significant number of low frequency codons in *H. salinarum* (**Figure 10**). The presence of low frequency codons in the rat AT1A GPCR, and more interestingly bR, provide evidence supporting our inclusion of low abundance codons in our degenerate strategy.

As shown in Table III, the degeneracy installed in our *mam2* sequence is not maximal. Aside from exclusion of rarely and

intermediately used codons, as mentioned above, codons proximal to the StyI and SphI restriction sites were not made degenerate to facilitate annealing of our oligonucleotides. Our degeneracy results in an experimentally feasible number of degenerate sequences, approximately 3^4, whereas maximal degeneracy would result in approximately 5^{10} sequences.

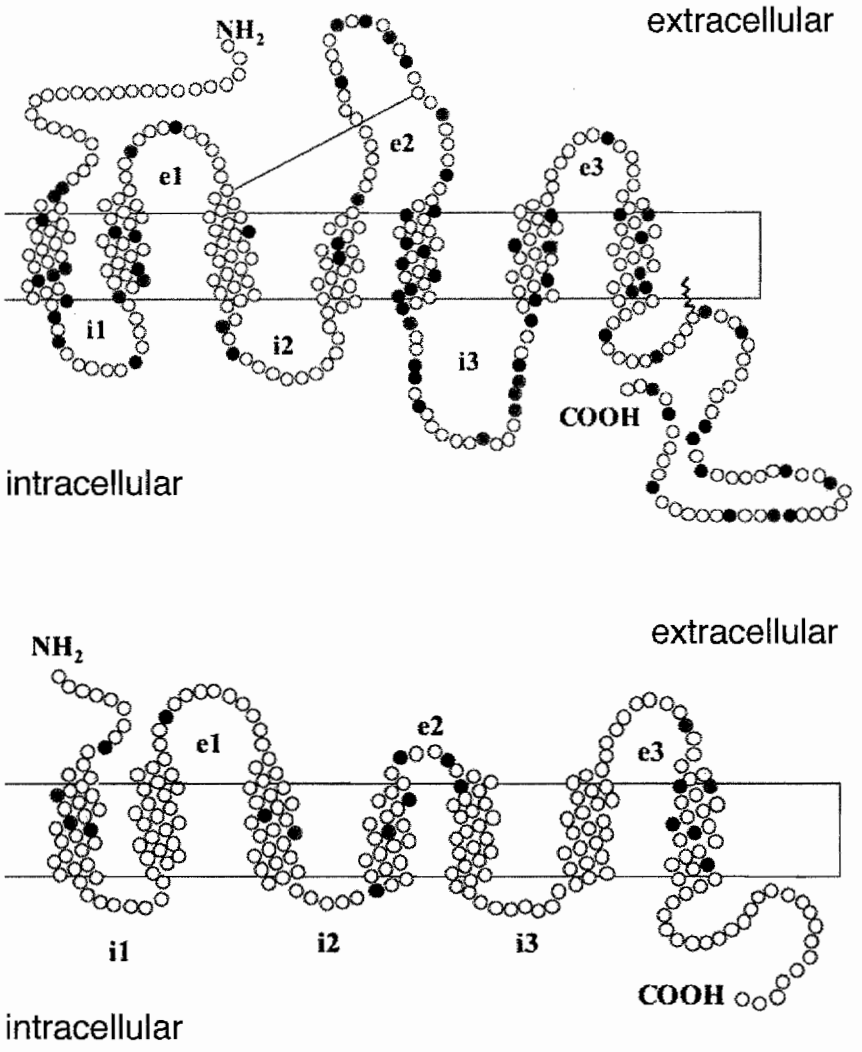


Figure 10. Codon Usage Abundance in the *angiotensin* and *bop* Gene Coding Regions

The amino acid secondary structures of the rat *angiotensin* type-1A (top) [88] and *bop* (bottom) gene coding regions are illustrated. The amino acids whose corresponding codons occur at low frequency in *H. salinarum* are shown as filled circles. Those codons used at frequencies of 0.1% and less are shown in red. Those codons used at frequencies between 0.5% and 0.1% are shown in black. The boxes indicate the boundaries of the membrane lipids. The amino- and carboxy-termini are labeled NH_2 and COOH, respectively. The extracellular and intracellular loop regions are labeled e1-e3 and i1-i3, respectively.

Conclusions

Understanding of signal transduction in GPCR is of great importance to the pharmaceutical and medical industries. Drugs targeting GPCR are responsible for roughly 200 billion dollars in annual sales and are associated with myriad diseases ranging from high blood-pressure to cancer. Research has been severely limited, however, due to the extremely low levels of GPCR expression. A novel chimeric protein containing intracellular GPCR components within a highly expressed membrane protein scaffold may provide insight into the mechanism of signal transduction. Here we fused the i3 loop of the Mam2 GPCR to the bacteriorhodopsin photoreceptor.

Bacteriorhodopsin provides an ideal scaffold for our fusion protein due to structural homology with GPCR, high levels of expression within *H. salinarum*, and ability to form highly resolvable crystals. The chosen GPCR to fuse to this scaffold was the Mam2 receptor due to the proven utility of yeast systems in GPCR studies, the association with a monomeric Gpa1, and existence of efficient *in vivo* signaling assays.

Experimental design of our bR:Mam2 fusion protein is limited considering the unknown structure of Mam2. However, through comparison with several structural predictions, the seven transmembrane domains of Mam2 were assigned resulting in identification of potential G-protein interacting domains, specifically the i3 loop. Although there is no evidence linking Gpa1 interaction with the Mam2 i3 loop in *Sz. pombe*, previous mutagenesis studies and comparison with other homologous pheromone receptors implicate the i3 loop as a critical component in signal transduction. Thus, we chose to

construct a bR:Mam2 i3 loop fusion protein in order to express high enough levels of protein to characterize interactions of Mam2 with Gpa1 during light-induced signal transduction.

The expression of chimeric proteins can only be achieved once successful cloning experiments have provided the genetic material to be expressed in a host system. Here, cloning of chimeric *bop* genes has been done for 34 degenerate sequences providing adequate material for transformation and expression in *H. salinarum*.

These 34 sequences were made degenerate in order to identify critical genetic factors regulating expression. Despite how widely studied bR has been in past several decades, critical determinants regulating such high expression in *H. salinarum* remains unknown. Successful expression of our bR:Mam2 fusions from known degenerate *bop:mam2* gene fusions will provide necessary information in determining genetic regulators of expression.

The cloning of these degenerate *bop* genes presents the first step toward expression of chimeric proteins in *H. salinarum* and is critical not only for creating a system to understand GPCR structurally and functionally, but to understand the genetic factors involved in protein expression. Complete cloning of 34 degenerate bR:Mam2 oligonucleotides has provided the material essential for protein expression, and transformation into *H. salinarum* may now be performed.

REFERENCES

1. Sarramegna, V. *et al.* **2003** *Cell. Mol. Life Sci.* 60 (8) 1529-1546
2. Grisshammer, R. and Tate, C. G. **2003** *Biochim. Biophys. Acta* 1610 (1) 1
3. Wess, J. **1998** *Pharmacol. Ther.* 80 (3) 231-264
4. Catapano, L. A. *et al.* **2007** *Biochim. Biophys. Acta-Biomembranes* 1768 (4) 976-993
5. Salazar, N. C. **2007** *Biochim. Biophys. Acta-Biomembranes* 1768 (4) 1006-1018
6. Metaye, T. *et al.* **2005** *Cell. Signalling* 17 (8) 917-928
7. Weir, C. J. **2007** *Anaesthesia & Intensive Care Medicine* 8 (10) 437-442
8. Palczewski, K. *et al.* **2000** *Science* 289, 739-745
9. Okada, T. *et al.* **2002** *Proc. Natl. Acad. Sci. USA* 99 (9) 5982-5987
10. Okada, T. *et al.* **2004** *J. Mol. Biol.* 342 (2) 571-583
11. Li, J. *et al.* **2004** *J. Mol. Biol.* 343 (5) 1409-1438
12. Kobilka, B. and Schertler, G. F. X. **2008** *Trends in Pharmacol. Sci.* 29 (2) 79-83
13. Baldwin, J. M. **1993** *EMBO J.* 12 (4) 1693-1703
14. Baldwin, J. M. **1997** *J. Mol. Biol.* 272 (1) 144-164
15. Kyte, J. and Doolittle, R. F. **1982** *J. Mol. Biol.* 157 (1) 105-132
16. Hjorth, S. A. *et al.* **1994** *J. Biol. Chem.* 269 (49) 30953-30959
17. Ohyama, K. *et al.* **1995** *Regulatory Peptides* 57 (2) 141-147
18. Chen, F. *et al.* **1992** *Mol. Endocrinol.* 6 (6) 914-919
19. Kobilka, B. K. and Deupi, X. **2007** *Trends in Pharmacol. Sci.* 28 (8) 397-406

20. Pan, H. *et al.* **2008** *Pharmacology. & Therapeutics* 117 (1) 141-161
21. Schoneberg, T. *et al.* **1999** *Mol. Cell. Endocrinology* 151 (1-2) 181-193
22. Beukers, M. W. *et al.* **2005** *Trends in Pharmacological Sciences* 26 (10) 533-539
23. Geiser, A. H. **2006** *Protein Sci.* 15, 1679-1690
24. Eyster, K. M. **1998** *Biochem. Pharmacol.* 55 (12) 1927-1938
25. Conklin, B. R. *et al.* **1993** *Nature* 363, 274-276
26. Liu, J. *et al.* **1995** *Proc. Natl. Acad. Sci. USA* 92 (25) 11642-11646
27. Cabrera-Vera, T. M. *et al.* **2003** *Endocrine Reviews* 24 (6) 765-781
28. Donnelly, D. and Findlay, J. B. C. **1994** *Cur. Op. Struc. Biol.* 4 (4) 582-589
29. Stenkamp, R. E. *et al.* **2002** *Biochim. Biophys. Acta-Biomembranes* 1565 (2) 168-182
30. Ballesteros, J. *et al.* **2001** *Curr. Opin. Drug Discov. Dev.* 4 (5) 561-574
31. Ballesteros, J. *et al.* **2001** *Mol. Pharmacol.* 60 (1) 1-19
32. Kenakin, T. **2002** *Nat. Rev. Drug Discov.* 1 (2) 103-110
33. Frauenfelder, H. *et al.* **1991** *Science* 254, 1598-1603
34. Kobilka, B. K. **2007** *Biochim. Biophys. Acta* 1768, 794-807
35. Hermans, E. **2003** *Pharmacology & Therapeutics* 99 (1) 25-44
36. Schertler, G. F. X. **1992** *Curr. Opin. Struct. Biol.* 2 (4) 534-544
37. Tate, C. G. and Grisshammer, R. **1996** *Trends in Biotechnology* 14 (11) 426-430
38. Jin, H. *et al.* **1998** *Biochim. Biophys. Acta* 1402, 165-170
39. Mirzadegan, T. *et al.* **2003** *Biochemistry* 42 (10) 2759-2767
40. Verrall, S. *et al.* **1997** *J. Biol. Chem.* 272 (11) 6898-6902
41. Liu, J. *et al.* **1996** *J. Biol. Chem.* 271 (15) 8772-8778

42. Kang, H. *et al.* **2005** *Biochem. and Biophys. Res. Comm.* 329 (2) 684-692
43. Strader, C. D. *et al.* **1994** *Annu. Rev. Biochem.* 63, 101-132
44. Teufel, M. *et al.* **1993** *EMBO J.* 12 (9) 3399-3408
45. Alblas, J. *et al.* **1995** *J. Biol. Chem.* 270 (15) 8944-8951
46. Parker, E. M. and Ross, E. M. **1991** *J. Biol. Chem.* 266 (15) 9987-9996
47. Gehret, A. U. *et al.* **2006** *J. Biol. Chem.* 281 (30) 20698-20714
48. Coria *et al.* **2001** *FEMS Microbiol. Lett.* 197 (1) 65-71
49. Shpakov, A. O. **2007** *J. of Evol. Biochem. and Phys.* 43 (1) 1-25
50. Turner, G. J. *et al.* **1999** *Protein Expr. Purif.* 17 (2) 324-338
51. Pausch, M. H. **1997** *Trends Biotechnol.* 15 (12) 487-494
52. Ladds, G. and Davey, J. **2004** *Curr. Opin. Drug Discov.* 7 (5) 683-691
53. Ladds, G. *et al.* **2003** *Mol. Microbiol.* 47 (3) 781-792
54. Kulkarni, R. D. *et al.* **2005** *Genome Biol.* 6 (3) R24
55. Hoffman, C. S. **2005** *Eukaryotic Cell* 4 (3) 495-503
56. Ladds, G. *et al.* **2007** *Cellular Signalling* 19, 103-113
57. Kitamura, K. and Shimoda, C. **1991** *EMBO J.* 10 (12) 3743-3751
58. Stefan, C. J. and Blumer, K. J. **1994** *Mol. Cell. Biology* 14 (5) 3339-3349
59. Stoecknius, W. and Rowen, R. **1967** *J. Cell Biol.* 34 (1) 365-393
60. Turner, G. J. **2008** Methods in Cell Biology 84, 479-515
61. Olsen, G.J. *et al.* **1997** *Cell* 89 (7) 991-994
62. Schafer, G. *et al.* **1999** *Microbiol. Mol. Biol. Rev.* 63 (3) 570-620

63. Winter-Vann, A. M. *et al.* **2001** Perspectives on Solid State NMR in Biology. 141-160
64. Oesterhelt, D. *et al.* **1974** *Methods in Enzymology* 31, 667-678
65. Turner, G. J. *et al.* **1999** *Protein Expr. Purif.* 17 (2) 312-323
66. Bartus, C. L. *et al.* **2003** *Biochim. Biophys. Acta* 1610, 109-123
67. Neutze, R. *et al.* **2002** *Biochim. Biophys. Acta* 1565, 144-167
68. Magyari, K. *et al.* **2006** *J. Photochem. and Photobiol. B: Biology* 85 (2) 140-144
69. Dunn, R. *et al.* **1981** *Proc. Natl. Acad. Sci. USA* 78 (11) 6744-6748
70. Gropp, F. *et al.* **1995** *Mol. Microbiol.* 16 (2) 357-364
71. Baliga, N. S. and DasSarma, S. **2000** *Mol. Microbiol.* 36 (5) 1175-1183
72. Gropp, R. *et al.* **1992** *Proc. Natl. Acad. Sci.* 89, 1204-1208
73. Ni, B. F. *et al.* **1990** *Gene* 90, 169-172
74. Krebs, M. P. *et al.* **1991** *Proc. Natl. Acad. Sci. USA* 88, 859-863
75. Heymann, J. A. *et al.* **1993** *Mol. Microbiol.* 7, 623-630
76. Turner, G. J. *et al.* **1993** *Biochemistry* 32, 1332-1337
77. Martinez, L. C. *et al.* **2002** *Proteins : Structure, Function, and Genetics* 48, 269-282
78. Ford, C. L. *et al.* **1991** *Protein Expression Purif.* 2, 95-107
79. Sohlemann, P. *et al.* **1997** *Naunyn-Schmiedeberg's Arch. Pharmacol.* 355, 150-160
80. Jaakola, V. P. *et al.* **2005** *Proteins* 60 (3) 412-423
81. DasSarma, S. *et al.* **1984** *Proc. Natl. Acad. Sci. USA* 81, 125-129

82. Tucker, J. and Grisshammer, R. **1996** *Biochem. J.* 317, 891-899

83. Gaborik, Z. *et al.* **1998** *FEBS Lett.* 428, 147-151

84. Engelman, D.M. *et al.* **1986** *Annu. Rev. Biophys. Chem.* 15, 321-353

85. Rost, B. *et al.* **1996** *Prot. Sci.* 7, 1704-1718

86. http://www.cbs.dtu.dk/services/GenomeAtlas/show-codon.php?KLSO=ASC&KLSK=ORGANISMSORT&kingdom=Archaea&GLgenus=Halobacterium&segmentid=Hspecies_NRC1_Main

87. Pfeifer, F. and Blaseio, U. **1990** *Nucleic Acids Res.* 18 (23) 6921-6925

88. Zitnay, C. and Siragy, H. **1998** *Mineral Electrolyte Metab.* 24, 362-370